

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-041512

(43)Date of publication of application : 08.02.2002

(51)Int.Cl.

G06F 17/27

(21)Application number : 2000-225911

(71)Applicant : OKI ELECTRIC IND CO LTD

(22)Date of filing : 26.07.2000

(72)Inventor : KITAMURA MIHOKO

MURATA TOSHIKI

SASAKI MIKI

SHIMOHATA SAYORI

FUKUI TAKASHI

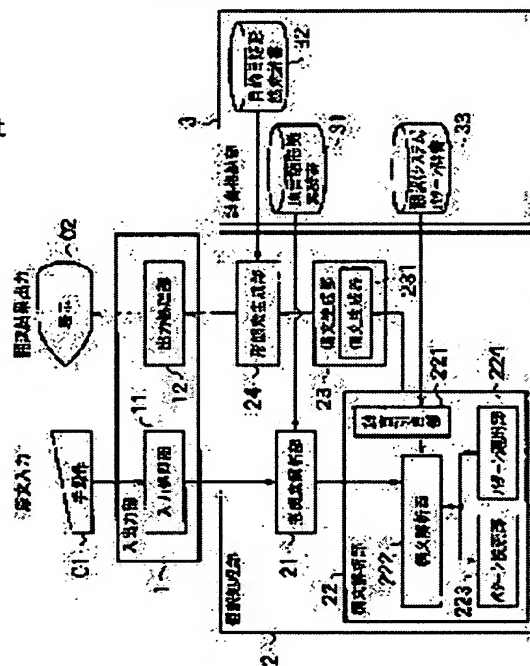
FUCHIGAMI MASAMUTSU

(54) DEVICE AND METHOD FOR NATURAL LANGUAGE PROCESSING

(57)Abstract:

PROBLEM TO BE SOLVED: To actualize a natural language processing which can provide an adequate natural language processing result (syntax analysis result and/or syntax generation result).

SOLUTION: This is a natural language processor which analyzes and/or generates syntax by using a natural language pattern having at least a pattern name and pattern constitution elements, and the device has at least a dictionary consulting means which extracts one or more natural language patterns as candidates for processing for syntax analysis and/or syntax generation, a pattern inspecting means which inspects whether or not the candidates for the natural language pattern is suitable to a tree structure, and a pattern applying means which applies the natural language pattern to the tree structure when the natural language pattern is suitable to the tree structure.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] Use a natural language pattern which has a pattern name and a pattern component at least, and it analyzes syntax and reaches. Or in natural-language-processing equipment which performs functor generation, it analyzes syntax and reaches from the above-mentioned natural language pattern currently beforehand prepared for a pattern dictionary. Or a dictionary handle stage which extracts one or more natural language patterns which serve as a candidate in processing of a functor generate time, Natural-language-processing equipment characterized by having a pattern inspection means to inspect whether a candidate's natural language pattern conforms to the tree structure, and a pattern application means to apply the natural language pattern to the tree structure when it suits.

[Claim 2] all or some which is beforehand prepared for a pattern dictionary of the above-mentioned natural language patterns — a pattern name — and — or natural-language-processing equipment according to claim 1 characterized by inspecting whether a pattern usage condition is given and the above-mentioned pattern inspection means conforms to the tree structure also with reference to a pattern usage condition of a candidate's natural language pattern about a pattern component.

[Claim 3] Natural-language-processing equipment according to claim 2 with which two or more natural language patterns with which only information about semantics differs are prepared for inside of two or more natural language patterns which have information about semantic constraint as one of the pattern usage conditions, and semantic information is characterized by determining the optimal tree structure through processing of the above-mentioned pattern inspection means and the above-mentioned pattern application means.

[Claim 4] Natural-language-processing equipment according to claim 1 to 3 characterized by having further a pattern evaluation means by which information on a priority to which it is given by it estimates the above-mentioned natural language pattern which serves as a candidate of application of the tree structure while information showing a priority over application is given to each above-mentioned natural language pattern currently beforehand prepared for a pattern dictionary.

[Claim 5] Natural-language-processing equipment according to claim 4 characterized by having raised a priority of a natural language pattern which has semantic constraint, and raising a priority of a natural language pattern which does not have semantic constraint between natural language patterns with which only existence of semantic constraint differs in a pattern name in a pattern component between natural language patterns with which only existence of semantic constraint differs.

[Claim 6] The above-mentioned pattern evaluation means is natural-language-processing equipment according to claim 4 or 5 characterized by removing except ** assistant [of a natural language pattern with highest priority] from a candidate when a candidate of two or more natural language patterns with which a pattern name and its pattern usage condition are the same with patterns, and information on a priority differs exists about application to the tree structure.

[Claim 7] The above-mentioned pattern evaluation means is natural-language-processing

equipment according to claim 4 to 6 characterized by usually excepting from a candidate of a natural language pattern of a low priority more relatively than a priority when a candidate of two or more natural language patterns with which a pattern name and a pattern component are the same with patterns, and information on a priority differs exists about application to the tree structure.

[Claim 8] The natural-language-processing equipment according to claim 4 to 7 characterized by to have a tree structure evaluation means evaluate superiority or inferiority of two or more tree structures, according to an evaluation allocation-of-marks method of also using information on a priority in a natural-language pattern which is the evaluation allocation-of-marks method that each tree structure can be evaluated, and was applied to subtree which is different by two or more tree structures when the tree structures which analyzed syntax, reached or were acquired by functor generation are two or more tree structures.

[Claim 9] The above-mentioned tree structure evaluation means is natural-language-processing equipment according to claim 8 characterized by making some of terminal symbols which constitute subtree which is different by two or more tree structures reflect in an evaluation allocation-of-marks method in addition to information on a priority in a natural language pattern applied to the tree structure.

[Claim 10] The above-mentioned tree structure evaluation means is natural-language-processing equipment according to claim 8 or 9 characterized by making a location of a node concerning a predetermined priority reflect in an evaluation allocation-of-marks method in addition to information on a priority in a natural language pattern applied to the tree structure.

[Claim 11] It is natural-language-processing equipment to claims 1-10 characterized by having a user registration means of a natural language pattern.

[Claim 12] Natural-language-processing equipment according to claim 11 characterized by giving a priority more than a natural language pattern of system registration to a natural language pattern of user registration.

[Claim 13] Natural-language-processing equipment according to claim 11 or 12 characterized by having a tree structure evaluation means to give top priority to the tree structure which has subtree to which a natural language pattern concerning user registration was applied in subtree which is different by two or more tree structures when the tree structures which analyzed syntax, reached or were acquired by functor generation are two or more tree structures.

[Claim 14] Use a natural language pattern which has a pattern name and a pattern component at least, and it analyzes syntax and reaches. Or in a natural-language-processing method of performing functor generation, it analyzes syntax and reaches from the above-mentioned natural language pattern currently beforehand prepared for a pattern dictionary. Or a consultation-of-a-dictionary production process which extracts one or more natural language patterns which serve as a candidate in processing of a functor generate time, A natural-language-processing method characterized by having a pattern inspection process which inspects whether a candidate's natural language pattern conforms to the tree structure, and a pattern application production process which applies the natural language pattern to the tree structure when it suits.

[Claim 15] all or some which is beforehand prepared for a pattern dictionary of the above-mentioned natural language patterns — a pattern name — and — or a natural-language-processing method according to claim 14 characterized by inspecting whether a pattern usage condition is given and the above-mentioned pattern inspection process conforms to the tree structure also with reference to a pattern usage condition of a candidate's natural language pattern about a pattern component.

[Claim 16] A natural-language-processing method according to claim 15 that two or more natural language patterns with which only information about semantics differs are prepared for inside of two or more natural language patterns which have information about semantic constraint as one of the pattern usage conditions, and semantic information is characterized by determining the optimal tree structure through processing of the above-mentioned pattern inspection process and the above-mentioned pattern application production process.

[Claim 17] A natural-language-processing method according to claim 14 to 16 characterized by having further a pattern evaluation production process that information on a priority to which it

is given by it estimates the above-mentioned natural language pattern which serves as a candidate of application of the tree structure while information showing a priority over application is given to each above-mentioned natural language pattern currently beforehand prepared for a pattern dictionary.

[Claim 18] A natural-language-processing method according to claim 17 characterized by having raised a priority of a natural language pattern which has semantic constraint, and raising a priority of a natural language pattern which does not have semantic constraint between natural language patterns with which only existence of semantic constraint differs in a pattern name in a pattern component between natural language patterns with which only existence of semantic constraint differs.

[Claim 19] The above-mentioned pattern evaluation production process is the natural-language-processing method according to claim 17 or 18 characterized by removing except ** assistant [of a natural language pattern with highest priority] from a candidate when a candidate of two or more natural language patterns with which a pattern name and its pattern usage condition are the same with patterns, and information on a priority differs exists about application to the tree structure.

[Claim 20] The above-mentioned pattern evaluation production process is the natural-language-processing method according to claim 17 to 19 characterized by usually excepting from a candidate of a natural language pattern of a low priority more relatively than a priority when a candidate of two or more natural language patterns with which a pattern name and a pattern component are the same with patterns, and information on a priority differs exists about application to the tree structure.

[Claim 21] The natural-language-processing method according to claim 17 to 20 characterized by to have a tree structure evaluation production process of evaluating superiority or inferiority of two or more tree structures, according to an evaluation allocation-of-marks method of also using information on a priority in a natural-language pattern which is the evaluation allocation-of-marks method that each tree structure can be evaluated, and was applied to subtree which is different by two or more tree structures when the tree structures which analyzed syntax, reached or were acquired by functor generation are two or more tree structures.

[Claim 22] The above-mentioned tree structure evaluation production process is the natural-language-processing method according to claim 21 characterized by making some of terminal symbols which constitute subtree which is different by two or more tree structures reflect in an evaluation allocation-of-marks method in addition to information on a priority in a natural language pattern applied to the tree structure.

[Claim 23] The above-mentioned tree structure evaluation production process is the natural-language-processing method according to claim 21 or 22 characterized by making a location of a node concerning a predetermined priority reflect in an evaluation allocation-of-marks method in addition to information on a priority in a natural language pattern applied to the tree structure.

[Claim 24] It is the natural-language-processing method to claims 14-23 characterized by having a user registration production process of a natural language pattern.

[Claim 25] A natural-language-processing method according to claim 24 characterized by giving a priority more than a natural language pattern of system registration to a natural language pattern of user registration.

[Claim 26] A natural-language-processing method according to claim 24 or 25 characterized by having a tree structure evaluation production process which gives top priority to the tree structure which has subtree to which a natural language pattern concerning user registration was applied in subtree which is different by two or more tree structures when the tree structures which analyzed syntax, reached or were acquired by functor generation are two or more tree structures.

[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[The technical field to which invention belongs] This invention can be applied to machine-translation equipment, the machine-translation method, etc. of performing a machine translation using a translation pattern, concerning natural-language-processing equipment and a method.

[0002]

[Description of the Prior Art] In the machine-translation equipment of a conventional type, the translation knowledge (a system dictionary and user's dictionary) which a system and a user can register was limited to the expression which machine-translation equipments, such as a collocation of a verb, such as "burning a hand", and a noun, have determined beforehand also by the word dictionary, the idiom which consists of two or more words, and the still more highly efficient thing. In order that the machine-translation equipment of a conventional type may analyze syntax in a source language sentence in the syntax and dictionary of the source language, may carry out language translation in a conversion dictionary and, as for this, may generate an object language sentence in the syntax and dictionary of object language, a system and a user are because a translation dictionary cannot be registered only in the format in accordance with the syntax of a system.

[0003] Furthermore, since the source language and object language are analyzed and generated according to an individual, they show below "[a noun phrase] based on intuition of a user as translation knowledge. "Following is [Noun phrase]. " (a noun phrase is a variable) (P1)

** -- the dictionary based on the pattern which consists of pairs of the source language [like] and object language cannot be registered, either.

[0004] The translation method based on the translation pattern indicated by JP,5-290082,A and translation equipment are one of those solved such a technical problem. Invention given [this] in an official report expresses translation knowledge within the limits of a context free language, and makes the syntax rule a pair by the pattern of the source language, and the pattern of object language. Moreover, the syntax rule made it possible to make speed of syntax analysis into practical use level by storing in the dictionary of a try mold. Furthermore, by having made the syntax rule into the pair by the pattern of the source language, and the pattern of object language, synchronous derivation was attained and language translation and functor generation came to require only easy processing.

[0005] The dictionary based on the translation pattern which consists of pairs of the source language like a pattern P1 and object language adapted to intuition of a user mentioned above by this can be registered now, if it is within the limits of a context free language, the translation knowledge of arbitration can be registered, and it came to be able to carry out translation processing.

[0006]

[Problem(s) to be Solved by the Invention] however -- choosing the solution of the minimum number of configurations (the number of patterns) about choosing the optimal solution from two or more solutions (for example, syntax-analysis result) in the above-mentioned official report -- only -- it is described. In this case, since the number of configurations serves as a parameter,

proper translation in consideration of the history of a component or the relation between components cannot be performed. Moreover, although the number of configurations (the number of patterns) is what specifies a priority, there is fear which is not right for a simple parameter.

[0007] Moreover, it is indicated by the above-mentioned official report that priority may be attached to a pattern. However, the clear method is not shown, but priority is given to a pattern how, and it is not indicated whether the optimal solution is caught.

[0008] This invention is made in consideration of the above point, and tends to offer the natural-language-processing equipment and the method of choosing a suitable natural-language-processing result.

[0009]

[Means for Solving the Problem] In order to solve this technical problem, the 1st this invention uses a natural language pattern which has a pattern name and a pattern component at least, and analyzes syntax and reaches. or natural-language-processing equipment which performs functor generation — setting — (1) — syntax analysis from the above-mentioned natural language pattern currently beforehand prepared for a pattern dictionary — and Or a dictionary handle stage which extracts one or more natural language patterns which serve as a candidate in processing of a functor generate time, (2) It is characterized by having a pattern inspection means to inspect whether a candidate's natural language pattern conforms to the tree structure, and a pattern application means to apply the natural language pattern to the tree structure when carrying out (3) adaptation.

[0010] The 2nd this invention uses a natural language pattern which has a pattern name and a pattern component at least, and analyzes syntax and reaches. or a natural-language-processing method of performing functor generation — setting — (1) — syntax analysis from the above-mentioned natural language pattern currently beforehand prepared for a pattern dictionary — and Or a consultation-of-a-dictionary production process which extracts one or more natural language patterns which serve as a candidate in processing of a functor generate time, (2) It is characterized by having a pattern inspection process which inspects whether a candidate's natural language pattern conforms to the tree structure, and a pattern application production process which applies the natural language pattern to the tree structure when carrying out (3) adaptation.

[0011]

[Embodiment of the Invention] (A) Explain the 1st operation gestalt which applied the natural-language-processing equipment and the method by this invention to machine-translation equipment and the machine-translation method in full detail below the 1st operation gestalt, referring to a drawing.

[0012] (A-1) The block diagram 1 of the 1st operation gestalt is a block diagram showing the functional configuration of the machine-translation equipment of the 1st operation gestalt. In addition, in practice, on information processors, such as a personal computer, loading of a processing program, the fixed data, etc. is carried out, and the machine-translation equipment of the 1st operation gestalt is built.

[0013] In drawing 1, the machine-translation equipment of the 1st operation gestalt consists of the I/O section 1, the translation processing section 2, and the dictionary storing section 3 greatly.

[0014] The I/O section 1 has the input-process section 11 which receives the input statement (source language sentence) which serves as a candidate for a translation from the input sections 01, such as a keyboard and file reading equipment, and the output-processing section 12 which outputs the translated result (object language sentence) to the output sections 02, such as a display, a printer, and file enclosure.

[0015] The translation processing section 2 consists of the morphological analysis section 21 which performs morphological analysis of an input statement, the syntax analyzer 22 which analyzes syntax of the source language using a translation pattern, the functor generation section 23 which generates functor by object language based on the result of syntax analysis of the source language, and the morphological generation section 24 which performs morphological generation of object language based on the result of functor generation.

[0016] A syntax analyzer 22 has the pattern Banking Inspection Department 223 which inspects whether the consultation-of-a-dictionary section 221 which performs pattern consultation of a dictionary, and the tree structure while building the pattern (source language pattern) looked up in a dictionary are suited, the pattern application section 224 applied to the tree structure while building a pattern by the inspection result, and the syntax analyzer 222 which calls these processing sections and builds the tree structure.

[0017] The dictionary storing section 3 consists of the source language morphological dictionary 31 in which the morphological information on the source language is stored, an object language morphological dictionary 32 in which the morphological information on object language is stored, and a translation pattern dictionary (system pattern dictionary) 33 in which the translation pattern which consists of a pair of the pattern of the source language and object language is stored.

[0018] (A-2) Explain actuation of the 1st operation gestalt, next translation actuation (the machine-translation method of the 1st operation gestalt) of the machine-translation equipment of the 1st operation gestalt. Here, drawing 2 is the flow chart which shows translation actuation of the machine-translation equipment of the 1st operation gestalt.

[0019] In addition, the machine-translation equipment of the 1st operation gestalt makes one sentence the input unit. Moreover, in the following explanation, actuation is explained to an example for the case where "The policeman arrests him." is inputted.

[0020] If a user inputs the above-mentioned input statement by the input section 01, the input-process section 11 will pass the morphological analysis section 21 of reception and the translation processing section 2 (step S21). The morphological analysis section 21 performs morphological analysis processing using the information on the source language morphological dictionary 31 (step S22).

[0021] A morphological analysis result is expressed by the tree structure which made the root node "Node", as shown in drawing 3. When it is a morpheme without two or more candidates, the standard form 3-1 of each morpheme and the morphological information 3-2, such as a part of speech and a change form, are given directly under a root node. On the other hand, when it is a morpheme with two or more candidates, each morphological candidate's information is given as a child node of the or node 3-3. For example, since "arrests" can take the part of speech of a verb and a noun, it is the morpheme which has two or more candidates, and as shown in drawing 3, each morphological candidate's information is given.

[0022] The syntax analyzer 22 to which such a morphological analysis result was given performs syntax-analysis processing to this result using the translation pattern dictionary 33 (step S23).

[0023] Drawing 4 takes out and shows some contents of storing of the translation pattern dictionary 33 (translation pattern).

[0024] In drawing 4, left-hand side expresses an English pattern, right-hand side expresses a Japanese pattern, and these English pattern and the Japanese pattern have become a pair. The pattern of each language is a [language name:pattern name pattern component].
since — it becomes.

[0025] A language name specifies English (en) and Japanese (ja).

[0026] As for the pattern name following a language name, the indicator in phrase structure rules, such as VP (verb phrase), NP (noun phrase), and N (noun), is applied.

[0027] a pattern component — two or more lists of a word, a variable, or a word and a variable — since — it becomes. A variable is described by [the numeric character:pattern name (it corresponds to the low order node of the tree structure) which is arbitration]. The numeric character portion of arbitration shows the correspondence relation between the source language used as a pair, and an object language pattern. In syntax analysis, a pattern can take an embedded structure by applying another pattern to a variable (a variable is canceled). Moreover, a word and a pattern name can have information with semantic detailed information etc. (history information) like the portion to which sign 4-1a and 4-1b were given. Furthermore, a word and a pattern name can variable-ize detailed information like the portion to which sign 4-7a was given, and also refer to the information for them.

[0028] As shown in the flow chart of drawing 5, syntax-analysis processing of the 1st operation

gestalt using such a translation pattern (source language pattern) consists of three processings, and mainly builds the tree structure to the bottom-up.

[0029] That is, checking that syntax analysis is not completed, three processings, pattern consultation-of-a-dictionary processing (step S52), pattern inspection processing (step S53), and pattern application processing (step S54), are repeated, and are performed (step S51).

[0030] The pattern consultation-of-a-dictionary processing (step S52) which is the 1st processing is processing which lengthens the translation pattern which has applicability next from a morphological analysis result and the result of the pattern application processing till then from the translation pattern dictionary 33. For example, the translation pattern (source language pattern) which attached the sign 4-4 of drawing 4 is lengthened from the translation pattern dictionary 33 from the morphological analysis result of "him:part-of-speech =prn" (sign 3-4) of drawing 3.

[0031] The pattern inspection processing (step S53) which is the 2nd processing is processing which inspects whether the translation pattern obtained as a result of consultation of a dictionary suits the tree structure under present construction for every tree structure.

[0032] The pattern application processing (step S54) which is the 3rd processing is processing which actually applies the pattern to the tree structure based on the tree structure and the translation pattern judged that suit as a result of inspection.

[0033] Above-mentioned pattern inspection processing and pattern application processing are concretely explained using drawing 6 and drawing 7.

[0034] The tree structure under present construction is drawing 6 (a), and suppose that the translation pattern to inspect was drawing 6 (c). The pattern name of a variable is the same at NP(s), and the detailed information which NP has is also in agreement by "semantic = people." Therefore, if the translation pattern shown in drawing 6 (c) suits the tree structure of drawing 6 (a), it will be judged by pattern inspection processing, the translation pattern is applied by pattern application processing, and the tree structure shown in drawing 7 is built.

[0035] On the other hand, since the detailed information (information on semantics) of NP differs even if the pattern name of a variable is the same at NP(s) when the tree structure under present construction is drawing 6 (b) and a translation pattern to inspect is drawing 6 (c), it is judged that the translation pattern concerned does not suit the tree structure by pattern inspection processing, and, naturally pattern application processing is not performed. In addition, in drawing 6 (b), "semantic != people" expresses "semantics is not a man." That is, "!=" expresses negation.

[0036] The syntax-analysis result of having applied the translation pattern as shown in drawing 4 comes to be shown in drawing 8 to input-statement "The policeman arrests him." In addition, to this input statement, the translation pattern 4-1 of drawing 4 and 4-3 to 4-7 are applied.

[0037] Such an syntax-analysis result is given to the functor generation section 23, and functor generation processing (step S24 of drawing 2) is performed by the functor generation section 23. In this functor generation processing, the tree structure of Japanese corresponding to an syntax-analysis result is acquired with reference to the translation pattern dictionary 33 using the pattern of the source language (English), and the pattern of the object language (Japanese) which is making the pair. Drawing 9 shows the functor generation result (tree structure) corresponding to the syntax-analysis result of drawing 8 mentioned above.

[0038] In addition, the translation pattern is the pair of a source language pattern and a translation pattern, and since the correspondence is a meaning, syntax-analysis processing and functor generation processing are performed mostly in parallel in practice. Moreover, functor generation may be performed to an syntax-analysis result by the method from old [which does not use a pattern].

[0039] The tree structure which the functor generation section 23 generated is given to the morphological generation section 24, and the morphological generation section 24 performs morphological generation processing (step S25). the Japanese word located in a termination node in a functor generation result in morphological generation processing — the order from the left — arranging — each word — the object language morphological dictionary 32 — using — preparing a verbal conjugated form **** — etc. — each word is operated orthopedically. The

output-processing section 12 carries out the display output of the morphological generation result (translation) by the receipt and the output section 02, or stores it in a file (step S26).

[0040] The translation result "a policeman arrests him" is obtained to input-statement "The policeman arrests him." mentioned above.

[0041] Here, the case where "The policeman arrests the deterioration." in which only an object differs from the above-mentioned input statement is inputted is considered.

[0042] In this case, since the semantics of the noun which comes to an object is except a man, the translation result "a policeman prevents aggravation" is obtained.

[0043] In addition, in the syntax analysis to this input statement, the translation pattern 4-2 of drawing 4, 4-3, and 4-5 to 4-8 are applied.

[0044] That is, it is judged with drawing 6 (b) suiting to the tree structure under construction over drawing 6 (a) and drawing 6 (b) which were mentioned above by adaptation decision with a pattern 4-8 and the pattern 4-2 with which the object (NP) by which 4-6 was inserted in the nest means except a man. Consequently, it is adopted, although the tree structure in which the information on a pattern 4-2 was incorporated is acquired by syntax analysis and this prevents as a translation of arrest.

[0045] As mentioned above, a different translation to the input statement which used the same verb etc. can be given by incorporating pattern usage conditions, such as "semantic = people" and "semantic != people", in the pattern.

[0046] (A-3) According to the 1st machine-translation equipment and method of an operation gestalt of an operation gestalt, the following effects can be done so. [of an effect 1st]

[0047] (a) It not only can make the translation pattern of the embedded structure by the variable, but it can carry out proper translation by conditions by giving pattern usage conditions, such as semantics, to a variable and a word.

[0048] Although the information equivalent to conditions was dropped on the official report which made reference by the term of a Prior art to English expression level like "claims \$ and \$" and "claim \$", in the case of this operation gestalt, it is not necessary to drop to English expression level. It is set to "claims [NP:number information = plurality]" and "claim [a NP:number information = unit]" with this operation gestalt.

[0049] (b) =! Description of pattern usage conditions, such as =, is easy, and can create a pattern easily. For example, if the semantics of an object "man" Becomes, if that is not right, like B, symbolic conventions are A and the format adapted to intuition of human being, and creation of a translation pattern is easy [a translation] for a translation.

[0050] "(c) — "I give these books." — functor — like — I give these a book" — two kinds of "I give these books" — existing — although — [— the former candidate may stop there being by registering a pattern with the conditions en:VP give [1:NP:semantics = people] and [2:NP]] Thus, the ambiguities of functor are also reducible by describing the pattern usage condition by semantics to a pattern.

[0051] (A-4) The deformation implementation gestalt above of the 1st operation gestalt explained using the example to which ambiguity does not exist in an syntax-analysis result. However, also when ambiguity exists, the 1st operation gestalt can be applied (when or node arises in an syntax-analysis result (refer to the 2nd operation gestalt)). For example, it may be generated, or functor generation may be performed making the syntax-analysis result which chooses any one of the child nodes of or node, and is unambiguous, and having ambiguity, and two or more translation results may be outputted.

[0052] It cannot be depended for the technical thought of the 1st operation gestalt on specific language, but it can be applied to the translation between any language.

[0053] Although the example of proper translation of the verb by the semantics of the noun located in the objective case was shown above, according to the semantics which the verb has in reverse, also when translating a noun properly, of course, it can apply, and there is only no objective case and proper translation by the nominative case is also possible. Proper translation of words other than a verb or a noun, such as proper translation of the preposition by the semantics of the word furthermore modified and the semantics of the noun phrase in a prepositional phrase, is also possible. Furthermore, also as for proper translation by pattern

usage conditions other than semantics, it is possible for a noun to determine the translation of a noun by the singular or the plural etc.

[0054] The processing about patterns, such as the pattern Banking Inspection Department, may be prepared in the exterior of syntax-analysis processing (refer to the 4th operation gestalt).

[0055] (B) Explain the 2nd operation gestalt which applied the 2nd natural-language-processing equipment and method by the operation gestalt, next this invention to machine-translation equipment and the machine-translation method in full detail, referring to a drawing.

[0056] (B-1) The block diagram 10 of the 2nd operation gestalt is a block diagram showing the functional configuration of the machine-translation equipment of the 2nd operation gestalt, and attaches and shows the same sign to the same [with drawing 1 concerning the 1st operation gestalt mentioned above], and a corresponding point.

[0057] In addition to the configuration of the 1st operation gestalt, the machine-translation equipment of the 2nd operation gestalt is equipped with the pattern evaluation section 225 in the syntax analyzer 22.

[0058] When two or more candidates' pattern exists, the pattern evaluation section 225 evaluates two or more tree structures which applied these patterns, and narrows down the tree structure.

[0059] In addition, the translation pattern stored in the translation pattern dictionary 33 also differs from the thing of the 1st operation gestalt. About this point, it clarifies by the term of operation.

[0060] (B-2) Also in the 2nd operation gestalt of ***** of the 2nd operation gestalt, the flow of the whole translation processing is as being shown in drawing 2 , and is performed in order of text input process, morphological analysis processing, syntax-analysis processing, functor generation processing, morphological generation processing, and translation result output processing.

[0061] Here, as for the 2nd operation gestalt, syntax-analysis processings differ as compared with the 1st operation gestalt.

[0062] Drawing 11 is a flow chart which shows syntax-analysis processing with the machine-translation equipment of the 2nd operation gestalt, and attaches and shows the same sign to the same [with drawing 5 concerning the 1st operation gestalt], and a correspondence processing step.

[0063] When it distinguishes whether two or more candidates' pattern exists corresponding to the tree structure node under construction (step S55) and two or more candidates' pattern exists after performing pattern consultation-of-a-dictionary processing (step S52), pattern inspection processing (step S53), and pattern application processing (step S54) one by one in the case of the 2nd operation gestalt, pattern evaluation processing (step S56) is performed.

[0064] Before explaining the concrete contents of processing of the newly added pattern evaluation processing, with reference to drawing 12 , the translation pattern stored in the translation pattern dictionary 33 of this 2nd operation gestalt is explained.

[0065] the -- two -- operation -- a gestalt -- a translation -- a pattern -- a head -- " -- [-- " -- a tail -- " --] -- " -- inserting -- having -- **** -- the -- one -- operation -- a gestalt -- being the same -- a pattern -- adding -- the head of a pattern -- "! There are a pattern 12-9 with which " was given, and a pattern 12-12 with which "*" was given to the head of a pattern. "!" and "*" are the special marks for expressing the priority of a pattern.

[0066] Drawing 13 is the example of the file which specifies the priority of a pattern, and the translation pattern dictionary 33 or the pattern evaluation section 22 stores it. Here, it is shown sequentially from the head of a file that a priority is high. Moreover, □ is ranking when the mark attaches nothing. Therefore, the priority in this example is "! It becomes the pattern with which " is given, a pattern without the mark of a priority, and the pattern with which "*" is given.

[0067] In drawing 12 , when the 2nd word is made into an adjective about the word which can take both a noun and an adjective after a definite article, and the English morphological train which the noun follows further, a pattern 12-9 is applied, and when the 2nd word is made into a noun, a pattern 12-10 is applied. As shown in drawing 12 , since the priority is raised from the pattern 12-10, a pattern 12-9 mentions details of operation later, but priority is given to a

pattern 12-9 if it is the case which can apply both patterns.

[0068] Moreover, in drawing 12, although both the pattern 12-11 and the pattern 12-12 are patterns about a sentence (S), the priority of the pattern 12-12 which is the more nearly general-purpose one where semantic constraint is not attached is low. Although details of operation are mentioned later, when the pattern 12-11 to which semantic constraint was given is not applied, the general-purpose pattern 12-12 will be adopted.

[0069] Pattern evaluation processing (step S56) estimates two or more tree structures to which two or more candidates' pattern was applied, using the priority of such a pattern.

[0070] Drawing 14 is a flow chart which shows the details of pattern evaluation processing. The pattern evaluation section 224 distinguishes whether there are two or more tree structure candidates whose detailed information (detailed information of a pattern name) of a root node corresponds first, or there is any tree structure candidate to whom the high priority was attached at the root node in a certain case (step S141). A high priority is relatively called high priority in a candidate here, and it is not necessarily "!" It is not limited to what "was given. And it removes except tree structure candidate ***** to which the high priority was attached, and its tree structure candidate from a candidate (step S142). Moreover, the pattern evaluation section 224 has two or more tree structure candidates whose components of a child node correspond, or, in a certain case, distinguishes whether there is any tree structure by which the low priority is given to the child node (step S143). Here, a low priority is relatively called low priority in a candidate, and it is not necessarily limited to what "*" was given. And if there is the tree structure to which the low priority is given, the tree structure will be removed from a candidate (step S144).

[0071] Next, input-statement "helium arrests the present deterioration." is explained to an example about the 2nd syntax-analysis processing of an operation gestalt, especially pattern evaluation processing. In the analysis to this input statement, the translation pattern shown in drawing 12 mentioned above is used suitably.

[0072] Drawing 15 (a) and (b) are the tree structures as a candidate (candidate for evaluation) who was able to do it by applying the pattern 12-9 shown in drawing 12, and 12-10 to the portion of "the presentdeterioration" in an input statement, respectively. " given to the tree structure shown in drawing 15 (a) at the translation pattern 12-9! The mark of "is attached. After termination of pattern application processing, if the tree structure with the equal and root node of the tree structure as shown in drawing 15 (a) and (b), and the equal word which constitutes the tree structure exists, these tree structures will be passed to pattern evaluation processing.

[0073] The contents of the priority symbolic file shown in drawing 13 are read beforehand, and pattern evaluation processing estimates application of a pattern according to it. If the tree structure which the mark to esteem without a mark attached temporarily exists, only when the detailed information which a root node has is equal, only a pattern with the highest mark is adopted and it deletes from the candidate of the tree structure in the middle of construction except it.

[0074] Since the priority is lower than the tree structure of drawing 15 (a) to which the mark which raises a priority was attached, the tree structure of drawing 15 (b) which the mark does not attach since the detailed information of the tree structure of drawing 15 (a) and the tree structure of drawing 15 (b) is equal in the case of the example of drawing 15 is deleted from the candidate of the tree structure.

[0075] On the other hand, drawing 16 is an example which has the tree structure to which mark "*" which lowers a priority from a pattern without a mark is attached in two or more candidates. It becomes conditions that the component of the child node of the tree structure is equal in evaluation of the mark which lowers a priority. drawing 16 -- (-- a --) -- being shown -- a thing -- a pattern -- 12 -- -- 11 -- applying -- having had -- a mark -- nothing -- the tree structure -- it is -- drawing 16 -- (-- b --) -- being shown -- a thing -- a pattern -- 12 -- -- 12 -- applying -- having had -- a priority -- lowering -- a mark -- " -- * -- " -- having been attached -- the tree structure -- it is -- these -- the tree structure -- a child node -- a

[0076] Drawing 17 shows the final syntax-analysis result of input-statement "helium arrests the

present deterioration.”

[0077] The Japanese tree structure corresponding to this syntax-analysis result is generated by functor generation processing, and, finally the translation result of “he prevents current aggravation” is outputted through morphological generation processing.

[0078] (B-3) According to the 2nd machine-translation equipment and method of an operation gestalt of an operation gestalt, while doing so the same effect as the 1st operation gestalt, also do an effect [like] so below further. [of an effect 2nd]

[0079] (a) When the tree structure candidate of right plurality exists in functor, there is a case where he wants to make a certain tree structure in it apply preferentially, and to translate it. Mark” which raises a priority with this 2nd operation gestalt! By giving ”to a pattern making it apply preferentially, this is realizable.

[0080] (b) In the middle of tree structure analysis, the conditions of a pattern are equal, when the pattern of two or more candidates who have a difference in the priority exists, by removing a pattern with a low priority from the candidate, analysis candidates can decrease in number and the processing times concerning syntax analysis or functor generation can be reduced.

[0081] (c) Only when only the pattern with which semantic constraint is when semantic constraint applies by registering the pattern which the semantic constraint to which the priority lowering mark was attached does not attach is applied and semantic constraint does not apply, it comes to apply to the pattern without semantic constraint which the priority lowering mark attached. The futility of the analysis of thereby always applying a pattern without semantic constraint can be lost, and the processing times concerning analysis are reduced. Furthermore, also when not applying semantic constraint, it does not become failure in syntax analysis, but the translation (it is general-purpose) which will seemingly be reasonable can be obtained.

[0082] (d) When applying to a pattern without semantic constraint by giving the mark which raises a priority to the pattern corresponding to the translation (it being general-purpose) which will seemingly be reasonable, it can translate not by the special translation with severe conditions but by the general-purpose translation.

[0083] (B-4) The mark which shows the deformation implementation gestalt priority of the 2nd operation gestalt may be what kind of mark (not only a mere mark but a numeric character, the class of parenthesis, a character string, etc. are included). Moreover, you may be not a mark but the information on other format.

[0084] Two or more priorities can also be prepared in order as two or more !, +, etc. are prepared and give priority to ! as a mark which raises a priority.

[0085] (C) Explain the 3rd operation gestalt which applied the 3rd natural-language-processing equipment and method by the operation gestalt, next this invention to machine-translation equipment and the machine-translation method in full detail, referring to a drawing.

[0086] (C-1) The block diagram 18 of the 3rd operation gestalt is a block diagram showing the functional configuration of the machine-translation equipment of the 3rd operation gestalt, and attaches and shows the same sign to the same [with drawing 10 concerning the 2nd operation gestalt mentioned above], and a corresponding point.

[0087] In addition to the configuration of the 2nd operation gestalt, the machine-translation equipment of the 3rd operation gestalt is equipped with the tree structure evaluation section 25.

[0088] When two or more candidates exist as an syntax-analysis result of the source language completed by syntax analysis, the information in which a pattern has which candidate is the most suitable estimates the tree structure evaluation section 25 (mark count), it chooses the optimal candidate (tree structure), and gives him to the functor generation section 23. Or two or more candidates are rearranged in order of an evaluation value, and it gives the functor generation section 23.

[0089] (C-2) Drawing 19 of the 3rd operation gestalt of operation is a flow chart explaining translation actuation of the machine-translation equipment of the 3rd operation gestalt, and attaches and shows the same sign to the same [with drawing 5 concerning the 1st operation gestalt], and a correspondence processing step.

[0090] In this 3rd operation gestalt, text input process (step S21), morphological analysis processing (step S22), syntax-analysis processing (step S23), tree structure evaluation

processing (step S27), functor generation processing (step S24), morphological generation processing (step S25), and translation result output processing (step S26) are performed by this order. That is, as compared with the 1st and 2nd operation gestalt, tree structure evaluation processing (step S27) is increasing.

[0091] Drawing 20 is a flow chart which shows tree structure evaluation processing. The tree structure evaluation section 25 will calculate an evaluating point by processing this to a bottom-up, if a syntax-analysis result is received. If a nonterminal node is taken out from the tip side of a branch (step S2001), it distinguishes whether it is or node (step S2002), if the tree structure evaluation section 25 is not or node, it will calculate the mark of the present node (step S2005), and if it is or node, after it makes an apex the mark of a node among the mark of a child node (step S2003), it will rearrange a node in order of mark (step S2004). Such processing is repeated until the nonterminal node which mark count has not ended is lost. In addition, the count method of concrete mark is mentioned later.

[0092] Hereafter, input-statement "It arrests an attention." is concretely explained to an example for tree structure evaluation processing.

[0093] Drawing 21 is as a result of [to this input statement] syntax analysis. Since two or nodes exist, this drawing 21 contains four syntax-analysis result candidates (the four tree structures are shown collectively). In drawing 21, the numeric character pulled out by the dotted line is the value calculated in tree structure evaluation processing, and is not included by the syntax-analysis result. In addition, the translation pattern dictionary 33 used by syntax-analysis processing presupposes that it is proportionate to the thing (drawing 12) of the 2nd operation gestalt. However, there are two kinds such as "!" and "+" as a priority raising mark, and the former presupposes that two steps of priority raising is expressed and the latter expresses one step of priority raising.

[0094] Drawing 22 shows the definition file of the count method of an evaluating point. The evaluating point count method definition file is stored in the tree structure evaluation section 25.

 [0095] The evaluating point count method definition file consists of the evaluation subject name field 221, an evaluation subfield 222, and the evaluating point field 223.

[0096] Evaluation subject names, such as terminal which means the mark "!", "+" and "*" of taking up and down of a priority, and "not only node" (general priority node) but the "number of the terminal symbols of the tree structure" concerning a pattern, and user (for example, registration to a morphological dictionary or a pattern dictionary) which means "the thing concerning user registration", are described by the evaluation subject name field 221.

[0097] Evaluation portions, such as total showing all the branches below a self-node (for example, or node) being evaluation portions and child showing the child node [directly under] of a self-node (for example, or node) being an evaluation portion, are described by the evaluation subfield 222.

[0098] The evaluating point is described by the evaluating point field 223. Here, the evaluating point is described by the formula (for example, $1000 - 100N$) according to the number (referred to as N) of evaluation criteria (for example, mark) with which an evaluation portion corresponds to what is total. In addition, mathematically, $1000 - 100N$ means $1000 - 100 \times N$. The evaluating point of immobilization is described to that whose evaluation portion is child.

[0099] The evaluating point count method definition file shown in drawing 22 evaluates the matter concerning user registration most, next evaluates the number of terminal symbols, and has after that the composition that the ranking of the priority of a pattern is evaluated. In addition, about the highest priority mark "!", it is making into the big evaluation wait whether to give the node directly under or node.

[0100] In the tree structure evaluation processing shown in drawing 20 mentioned above, mark count of step S2005 is made according to the contents of a convention of the evaluating point count method definition file.

[0101] Drawing 21 explains the count method of a concrete evaluating point.

[0102] The "+" mark is given, and since the number of nonterminal symbols is one of only attention, $+N$ of a node 21-1 The record of 225 and 230 in the definition file of drawing 22 is applied, and an evaluating point turns into 90000001 points. On the other hand, a node 21-2 The

"+" mark is not given, and since the number of nonterminal symbols is one of only attention, the record of 225 in the definition file of drawing 22 is applied, and an evaluating point turns into 90 million points. Consequently, in the following or node 21-3, the evaluating point of the node 21-1 of the higher one is adopted.

[0103] Moreover, since it is 1 Padding and the number of terminal symbols (the number of termination nodes) is also one of the number of child nodes, "arrests a attention" VP of a node 21-4 The records 225 and 228 of a file are applied and an evaluating point turns into 90000900 points. On the other hand, VP of a node 21-5 Since three pieces and the number of child nodes have [the number of terminal symbols (the number of termination nodes)] four pieces and the "+" mark in one tip side of a branch, the records 225, 228, and 2210 of a file are applied, and an evaluating point turns into 7000601 points. Consequently, the mark of a node 21-4 are adopted in the following or node 21-6.

[0104] After count of an evaluating point is completed as mentioned above, the sequence of a node is changed together with the order of mark. Thus, the tree structure after rearranging is shown in drawing 23 .

[0105] Functor generation processing generates the tree structure in Japanese (object language) corresponding to this tree structure. Temporarily, if or node is in the English tree structure, only the Japanese node corresponding to the English node located in the very head by the child node of or node will be chosen, and generation processing will be performed. Then, morphological generation processing is performed, consequently the translation result "it lengthens interest" is outputted.

[0106] In addition, temporarily, in wanting a user to output two or more translation result candidates, it performs functor generation and morphological generation processing using the Japanese tree structure corresponding to all the child nodes of or node. The translation result in this case becomes "it (| cautions which lengthen interest are prevented)." In this case, it is shown that former one is evaluating the system.

[0107] (C-3) According to the 3rd machine-translation equipment and method of an operation gestalt of an operation gestalt, the following effects can be done so while doing so the same effect as the 1st and 2nd operation gestalten. [of an effect 3rd]

[0108] It is alike and has the following effects more.

[0109] (a) Finer proper translation and proper translation by the debt by syntax analysis become possible by evaluating by the whole tree structure using the various priority marks given to the pattern and the detailed information of a pattern.

[0110] (b) According to this 3rd operation gestalt, the translation result of having followed priority can be expected and it is easy to predict a translation result by deciding the priority of various priority marks or pattern detailed information.

[0111] (c) The same formula can also estimate conditions other than priority marks, such as the number of termination patterns, to coincidence.

[0112] (d) For example, in the analysis result of "--today--", when only the candidate of a noun exists, it translates as a noun, but when saying that he wants to translate as an adverb when the candidate of both a noun and an adverb exists, the priority of many parts of speech can be controlled by giving the mark which raises a priority to the pattern of a part of speech (adverb in this case) to give priority to.

[0113] (C-4) The priority mark which controls the count method of an evaluating point used with the operation gestalt of the deformation implementation gestalt 3rd of the 3rd operation gestalt may use together the priority mark of the 2nd operation gestalt, and you may make it differ. What can reduce candidates by pattern evaluation processing by grant of a mark in the case of the former will be reduced, and two steps of priority processings will be performed in except [its] by one mark of deciding a priority by mark count. The above-mentioned explanation is conscious of such a case.

[0114] However it may give the mark which specifies the count method of an evaluating point, no matter it may be what class, it is not cared about. Moreover, it is also possible to give two or more kinds of marks to coincidence.

[0115] The formula of mark and allocation of marks of each evaluation criteria may be any

equations or any allocation of marks. Moreover, you may enable it to change easily. Carrying out mark count which took the level of the tree structure into consideration can also make higher than the priority of the number of the marks of the whole tree structure priority of the mark given to the node directly under or node.

[0116] A priority can be decided according to not only the number of termination nodes but other tree structures, or the feature of a pattern. For example, a fixed principle can be given to the output of a translation result as the severe pattern of conditions is given [priority] and translated even when two or more patterns apply to conditions whenever the thing which has the few number of the semantics of the noun phrase of the object of the 1st operation gestalt gives a mark so that priority may be given.

[0117] (D) Explain the 4th operation gestalt which applied the 4th natural-language-processing equipment and method by the operation gestalt, next this invention to machine-translation equipment and the machine-translation method in full detail, referring to a drawing.

[0118] (D-1) The block diagram 24 of the 4th operation gestalt is a block diagram showing the functional configuration of the machine-translation equipment of the 4th operation gestalt, and attaches and shows the same sign to the same [with drawing 18 concerning the 3rd operation gestalt], and a corresponding point.

[0119] The machine-translation equipment of the 4th operation gestalt is equipped with the portions of the pattern Banking Inspection Department 251, the pattern application section 252, and the pattern evaluation section 253 as the pattern processing section 26 which became independent to the exterior of a syntax analyzer 22, and the functor generation machine 231 in the functor generation section 23 enables it to use it.

[0120] That is, not only the time of syntax analysis but the functor generate time enables it to perform inspection of a pattern, application, and evaluation.

[0121] In addition, two or more things also have an object language pattern corresponding to the same source language pattern as a translation pattern. Moreover, although drawing 4 does not show, naturally a pattern usage condition etc. can also be given to an object language pattern. This 4th operation gestalt is premised on such a thing.

[0122] (D-2) Different actuation from the operation gestalt which the operation gestalt of the actuation 4th of the 4th operation gestalt mentioned already is functor generation processing. So, below, the characteristic processing by functor generation processing of the 4th operation gestalt is explained, referring to the flow chart of drawing 25.

[0123] By the conventional method, since the Japanese (object language) tree structure is also built almost in parallel to construction of the English (source language) tree structure (replacement of the tree structure using a simple pattern), special functor generation processing is not performed. Therefore, it is undetectable even if a gap of a pattern usage condition arises in the Japanese tree structure.

[0124] In the tree structure of Japanese built almost in parallel to construction of the English tree structure, this 4th operation gestalt tends to cancel two or more candidates by performing pattern inspection processing, pattern application processing, and pattern evaluation processing, when two or more candidate node occurs. As opposed to the English tree structure, drawing 25 is performed to the tree structure of temporary Japanese (object language) which acquired the Japanese tree structure by the conventional functor generation processing (simple replacement using a pattern).

[0125] First, one-piece ejection (step S2501) and it distinguish whether it is or node for a nonterminal symbol from the tree structure with the functor generation machine 231 temporary from a root node to a top-down (step S2502).

[0126] If there is an or node, the pattern Banking Inspection Department 223 will be made to perform delivery, pattern inspection processing, and pattern application processing for the tree structure of parents including the pattern and self-node of a child node of or node (step S2503). When pattern inspection processing is completed, according to the inspection result, pattern application processing by the pattern application section 224 is performed (step S2504). Then, it distinguishes whether or node was canceled by pattern application processing (step S2505), and if not canceled, the pattern evaluation section 225 is made to perform pattern evaluation

processing (step S2507).

[0127] When the taken-out nonterminal symbol is not or node, the node is made to apply as it is (step S2506).

[0128] The tree structure of a repeat and Japanese (object language) is completed until it becomes impossible for a nonterminal symbol to take out such processing (generation).

[0129] In addition, pattern inspection processing, pattern application processing, and pattern evaluation processing are the almost same processings as the operation gestalt mentioned already.

[0130] Hereafter, functor generation processing with the 4th operation gestalt is explained to an example for an input statement "the intellectual woman."

[0131] Drawing 26 (a) shows the syntax-analysis result (tree structure in English) of this input statement, and drawing 26 (b) shows the tree structure in Japanese corresponding to it. The Japanese tree structure of drawing 26 (b) is the temporary tree structure made from the correspondence result of the English tree structure. Processing shown in drawing 25 mentioned above is performed to this temporary tree structure.

[0132] In the tree structure shown in drawing 26 (b), or node is contained and pattern inspection processing and pattern application processing are performed.

[0133] If it is the case where the translation pattern as shown in drawing 27 (b1) is described here and is a case as the tree structure of the parents who present pattern inspection processing and pattern application processing shows drawing 27 (a), the result that a pattern suits will be obtained at pattern inspection, and the Japanese pattern of drawing 27 (b1) will be applied. Consequently, it is canceled and or node progresses to the following node.

[0134] However, a translation pattern as shown in drawing 27 (b2) is also the case where two kinds of Japanese patterns are described corresponding to the English pattern, and if it is the case where both patterns are applicable, pattern evaluation processing will be performed.

[0135] Since the priority mark is given to two kinds of Japanese patterns, it is narrowed down to one piece here. Or priority is given.

[0136] Drawing 28 shows the functor generation processing result at the time of giving priority to the above-mentioned input statement. This result is passed to the morphological generation section and the translation result of "an intellectual woman" or "the woman who requires intellectual female | intellect" is obtained.

[0137] (D-3) the effect of the operation gestalt which was mentioned already according to the 4th machine-translation equipment and method of an operation gestalt of an operation gestalt — in addition, the following effects can be done so. [of an effect 4th]

[0138] (a) It can translate properly not only using the information about the source language but using the information on object language.

[0139] (b) When the bidirectional translation in the above-mentioned translation pattern dictionary is carried out, by giving a priority to both the source language and an object language pattern, it is an analysis and generation side and each priority can be used again.

[0140] (D-4) Although the deformation implementation gestalt above of the 4th operation gestalt explained the method similar to the 2nd operation gestalt of reducing two or more candidates to a pattern generate time, a syntax analyzer is very good in a method similar to the 3rd operation gestalt of asking for all candidates and carrying out mark count of a functor spanning tree after that. Moreover, it is very good in both methods.

[0141] When the reverse (that is, translation to the source language from object language) of a translation of object language is considered from the source language, the priority given to the pattern of object language can be used at the time of analysis. In that case, the priority given to the pattern by the side of the source language can be used for a generate time.

[0142] (E) Explain the 5th operation gestalt which applied the 5th natural-language-processing equipment and method by the operation gestalt, next this invention to machine-translation equipment and the machine-translation method in full detail, referring to a drawing.

[0143] (E-1) The block diagram 29 of the 5th operation gestalt is a block diagram showing the functional configuration of the machine-translation equipment of the 4th operation gestalt, and attaches and shows the same sign to the same [with drawing 24 concerning the 4th operation

gestalt], and a corresponding point.

[0144] This 5th operation gestalt is equipped with the user pattern dictionary 34 which equips the I/O section 1 with the user's dictionary registration processing section 13 which receives the translation pattern which a user inputs, and stores that translation pattern in it at the dictionary storing section 3. That is, a user registration (deletion) function is carried.

[0145] Therefore, a syntax analyzer 22 etc. will refer to both the system pattern dictionary 33 and the user pattern dictionary 34.

[0146] (E-2) Since the thing concerning the user registration of the 5th operation gestalt of operation is a translation pattern, although a know how is needed, when direct registration of detailed information, such as conditions, is equipped with a user interface etc., it can register by the same processing as registration of the element of a word (morpheme) and others. However, the following points differ.

[0147] "Following is "[a noun phrase] is shown below for [noun phrase]." The case where a user registers "is explained.

[0148] The translation pattern in this case is [en:S [following is [1:NP]].

[ja:S [1:NP] is shown below.] It becomes.

[0149] It is [en:S [following is [1:NP:semantics, when the word located in a noun phrase is except a man and you want to apply this pattern temporarily! = man] .] and [ja:S [1:NP] are shown below.] Thus, semantic constraint is also described and registered. In this case, the user's dictionary registration processing section 13 displays the message which asks whether conditioning is performed or not in NP etc., the example of a setting of conditions, etc., and you may make it it not only to await the input of the arbitration by the user, but incorporate conditions.

[0150] Moreover, the highest priority raising mark will always be added to the pattern concerning user registration, and will be registered into it.

[0151] Furthermore, as mentioned above, in the evaluating point count method definition file (the 3rd operation gestalt is started) of drawing 22 , what he is trying to give the highest evaluating point to the record 221 of an evaluation subject name called user, and starts the pattern of a user pattern dictionary at the time of evaluation of the tree structure is given top priority, and the translation result is obtained.

[0152] Here, user shows a user pattern and child shows directly under a child node. A record 22-1 means ** which gives an apex (100 million points) to the user pattern directly under or node. Thereby, the pattern with which the user registered the user pattern by a priority becoming high to all candidates surely comes to be reflected in a translation result.

[0153] (E-3) the effect of the operation gestalt which was mentioned already according to the 5th machine-translation equipment and method of an operation gestalt of an operation gestalt — in addition, the following effects can be done so. [of an effect 5th]

[0154] (a) Semantics and other conditions can also be given at a variable to the translation pattern dictionary which a user draws up. Being able to perform registration by the accepted pattern by this, there are few patterns to register and it ends.

[0155] (b) It is also possible to give a priority mark for every user pattern, and control of the translation by the user or a translation is also possible.

[0156] (c) Since the evaluating point concerning a user pattern is also performed by the same mark count as the count to other patterns, it can respond also to modification of the priority (priority mark) of a user pattern dictionary etc. easily.

[0157] (d) A user can also obtain the translation result which does not use a user pattern to the translation and coincidence using a user pattern by outputting all candidate translations.

[0158] (E-4) With the operation gestalt of the deformation implementation gestalt 5th of the 5th operation gestalt, although the user pattern dictionary was made top priority, giving priority to the number of the minimum nodes, next giving priority to a user pattern etc. can change priority suitably.

[0159] Two or more user pattern dictionaries can be drawn up, and priority can be attached to each dictionary.

[0160] The mark which raises or lowers a priority also in a user pattern dictionary can be given.

(In that case, it becomes priority attachment in a user pattern dictionary.)

Creation of a user pattern dictionary can be made easy by setting in the user pattern dictionary input-process section, and a user's not doing the direct input of the translation patterns, such as [en:—] and [ja:—], but preparing the user interface which inputs a pattern name, a word, and a variable simply.

[0161] In the pattern evaluation section, although the 5th operation gestalt showed the method which gives priority to and translates a user pattern dictionary by evaluating a user pattern by the tree structure evaluation section, when a user pattern and the other candidate exist, a user pattern can be applied and the other candidate can be realized also by the method of not applying (refer to the 2nd operation gestalt).

(F) The candidate for application of the technical thought of other operation gestalt this inventions is limited to neither machine-translation equipment nor a method, but can be applied to the natural-language-processing equipment and the method of analyzing syntax using a natural language pattern, and the natural-language-processing equipment and the method of performing functor generation using a natural language pattern.

[0162] In this case, syntax analysis may be performed using a natural language pattern, functor generation may be performed by the method of not using a natural language pattern, and you may be that reverse. Moreover, only syntax analysis is needed and this invention can be applied to the equipment which does not perform functor generation, the equipment of the reverse, etc.

[0163]

[Effect of the Invention] As mentioned above, according to this invention, the natural-language-processing equipment and the method of offering a suitable natural-language-processing result are realizable.

[Translation done.]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2002-41512

(P2002-41512A)

(43) 公開日 平成14年2月8日 (2002.2.8)

(51) Int.Cl.⁷

G 0 6 F 17/27

識別記号

F I

G 0 6 F 17/27

テーマコード(参考)

J 5 B 0 9 1

審査請求 未請求 請求項の数26 O L (全 25 頁)

(21) 出願番号 特願2000-225911(P2000-225911)

(22) 出願日 平成12年7月26日 (2000.7.26)

(71) 出願人 000000295

沖電気工業株式会社

東京都港区虎ノ門1丁目7番12号

(72) 発明者 北村 美穂子

東京都港区虎ノ門1丁目7番12号 沖電気
工業株式会社内

(72) 発明者 村田 稔樹

東京都港区虎ノ門1丁目7番12号 沖電気
工業株式会社内

(74) 代理人 100090620

弁理士 工藤 宜幸

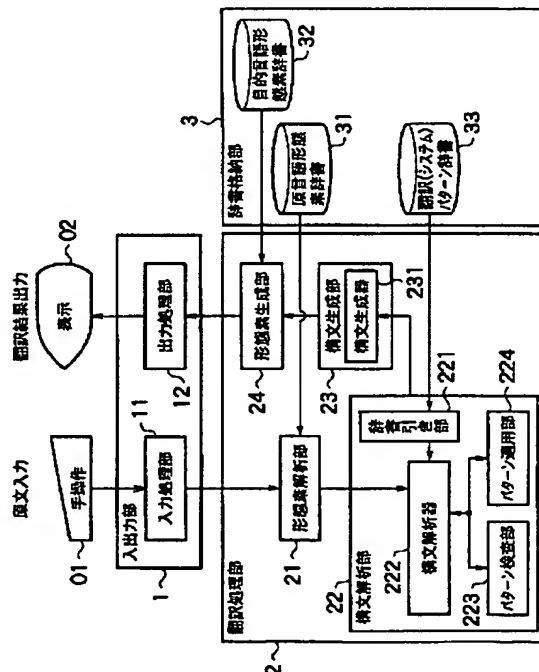
最終頁に続く

(54) 【発明の名称】 自然言語処理装置及び方法

(57) 【要約】

【課題】 適切な自然言語処理結果（構文解析結果及び又は構文生成結果）を提供し得る自然言語処理装置を実現する。

【解決手段】 本発明は、少なくともパターン名及びパターン構成要素を有する自然言語パターンを利用して、構文解析、及び又は、構文生成を行う自然言語処理装置に関する。そして、予めパターン辞書に用意されている上記自然言語パターンから、構文解析、及び又は、構文生成時の処理での候補となる1以上の自然言語パターンを抽出する辞書引き手段と、候補の自然言語パターンが木構造に適合しているか否かを検査するパターン検査手段と、適合する場合に、その自然言語パターンを木構造に適用するパターン適用手段とを少なくとも有することを特徴とする。



【特許請求の範囲】

【請求項 1】 少なくともパターン名及びパターン構成要素を有する自然言語パターンを利用して、構文解析、及び又は、構文生成を行う自然言語処理装置において、予めパターン辞書に用意されている上記自然言語パターンから、構文解析、及び又は、構文生成時の処理での候補となる 1 以上の自然言語パターンを抽出する辞書引き手段と、候補の自然言語パターンが木構造に適合しているか否かを検査するパターン検査手段と、適合する場合に、その自然言語パターンを木構造に適用するパターン適用手段とを有することを特徴とする自然言語処理装置。

【請求項 2】 予めパターン辞書に用意されている全て又は一部の上記自然言語パターンについて、パターン名及び又はパターン構成要素について、パターン適用条件を持たせておき、上記パターン検査手段が、候補の自然言語パターンのパターン適用条件をも参照して木構造に適合しているか否かを検査することを特徴とする請求項 1 に記載の自然言語処理装置。

【請求項 3】 パターン適用条件の 1 つとして意味条件に関する情報を有する複数の自然言語パターンのうちに、意味に関する情報のみが異なる複数の自然言語パターンを用意しておき、上記パターン検査手段及び上記パターン適用手段の処理を通じて、意味情報が最適な木構造を決定することを特徴とする請求項 2 に記載の自然言語処理装置。

【請求項 4】 予めパターン辞書に用意されている上記各自然言語パターンには、適用に対する優先度を表す情報が付与されていると共に、木構造の適用の候補となっている上記自然言語パターンを、それに付与されている優先度の情報により評価するパターン評価手段をさらに有することを特徴とする請求項 1～3 のいずれかに記載の自然言語処理装置。

【請求項 5】 パターン構成要素において、意味条件の有無だけが異なる自然言語パターンの間では、意味条件を有する自然言語パターンの優先度を高め、パターン名において、意味条件の有無だけが異なる自然言語パターンの間では、意味条件を持たない自然言語パターンの優先度を高めたことを特徴とする請求項 4 に記載の自然言語処理装置。

【請求項 6】 上記パターン評価手段は、木構造への適用に関し、パターン名及びそのパターン適用条件が同じであって優先度の情報が異なる複数の自然言語パターンの候補が存在した場合に、優先度が最も高い自然言語パターンの候補以外を候補から除くことを特徴とする請求項 4 又は 5 に記載の自然言語処理装置。

【請求項 7】 上記パターン評価手段は、木構造への適用に関し、パターン名及びパターン構成要素が同じであ

って優先度の情報が異なる複数の自然言語パターンの候補が存在した場合に、通常優先度より相対的に低い優先度の自然言語パターンの候補から除外することを特徴とする請求項 4～6 のいずれかに記載の自然言語処理装置。

【請求項 8】 構文解析、及び又は、構文生成で得られた木構造が複数の木構造である場合に、各木構造を評価し得る評価配点方法であって、複数の木構造で異なっている部分木に適用された自然言語パターンでの優先度の情報をも利用している評価配点方法に従って、複数の木構造の優劣を評価する木構造評価手段を有することを特徴とする請求項 4～7 のいずれかに記載の自然言語処理装置。

【請求項 9】 上記木構造評価手段は、木構造に適用された自然言語パターンでの優先度の情報に加え、複数の木構造で異なっている部分木を構成する終端記号の多少をも評価配点方法に反映させていることを特徴とする請求項 8 に記載の自然言語処理装置。

【請求項 10】 上記木構造評価手段は、木構造に適用された自然言語パターンでの優先度の情報に加え、所定の優先度に係るノードの位置をも評価配点方法に反映させていることを特徴とする請求項 8 又は 9 に記載の自然言語処理装置。

【請求項 11】 自然言語パターンのユーザ登録手段を有することを特徴とする請求項 1～10 に自然言語処理装置。

【請求項 12】 ユーザ登録の自然言語パターンにシステム登録の自然言語パターン以上の優先度を付与することを特徴とする請求項 11 に記載の自然言語処理装置。

【請求項 13】 構文解析、及び又は、構文生成で得られた木構造が複数の木構造である場合に、複数の木構造で異なっている部分木において、ユーザ登録に係る自然言語パターンを適用された部分木を有する木構造を最優先する木構造評価手段を有することを特徴とする請求項 11 又は 12 に記載の自然言語処理装置。

【請求項 14】 少なくともパターン名及びパターン構成要素を有する自然言語パターンを利用して、構文解析、及び又は、構文生成を行う自然言語処理方法において、

予めパターン辞書に用意されている上記自然言語パターンから、構文解析、及び又は、構文生成時の処理での候補となる 1 以上の自然言語パターンを抽出する辞書引き工程と、

候補の自然言語パターンが木構造に適合しているか否かを検査するパターン検査工程と、適合する場合に、その自然言語パターンを木構造に適用するパターン適用工程とを有することを特徴とする自然言語処理方法。

【請求項 15】 予めパターン辞書に用意されている全て又は一部の上記自然言語パターンについて、パターン

名及び又はパターン構成要素について、パターン適用条件を持たせておき、

上記パターン検査工程が、候補の自然言語パターンのパターン適用条件をも参照して木構造に適合しているか否かを検査することを特徴とする請求項 14 に記載の自然言語処理方法。

【請求項 16】 パターン適用条件の 1 つとして意味条件に関する情報を有する複数の自然言語パターンのうちに、意味に関する情報のみが異なる複数の自然言語パターンを用意しておき、

上記パターン検査工程及び上記パターン適用工程の処理を通じて、意味情報が最適な木構造を決定することを特徴とする請求項 15 に記載の自然言語処理方法。

【請求項 17】 予めパターン辞書に用意されている上記各自然言語パターンには、適用に対する優先度を表す情報が付与されていると共に、

木構造の適用の候補となっている上記自然言語パターンを、それに付与されている優先度の情報により評価するパターン評価工程をさらに有することを特徴とする請求項 14～16 のいずれかに記載の自然言語処理方法。

【請求項 18】 パターン構成要素において、意味条件の有無だけが異なる自然言語パターンの間では、意味条件を有する自然言語パターンの優先度を高め、パターン名において、意味条件の有無だけが異なる自然言語パターンの間では、意味条件を持たない自然言語パターンの優先度を高めたことを特徴とする請求項 17 に記載の自然言語処理方法。

【請求項 19】 上記パターン評価工程は、木構造への適用に関し、パターン名及びそのパターン適用条件が同じであって優先度の情報が異なる複数の自然言語パターンの候補が存在した場合に、優先度が最も高い自然言語パターンの候補以外を候補から除くことを特徴とする請求項 17 又は 18 に記載の自然言語処理方法。

【請求項 20】 上記パターン評価工程は、木構造への適用に関し、パターン名及びパターン構成要素が同じであって優先度の情報が異なる複数の自然言語パターンの候補が存在した場合に、通常優先度より相対的に低い優先度の自然言語パターンの候補から除外することを特徴とする請求項 17～19 のいずれかに記載の自然言語処理方法。

【請求項 21】 構文解析、及び又は、構文生成で得られた木構造が複数の木構造である場合に、各木構造を評価し得る評価配点方法であって、複数の木構造で異なっている部分木に適用された自然言語パターンでの優先度の情報をも利用している評価配点方法に従って、複数の木構造の優劣を評価する木構造評価工程を有することを特徴とする請求項 17～20 のいずれかに記載の自然言語処理方法。

【請求項 22】 上記木構造評価工程は、木構造に適用された自然言語パターンでの優先度の情報に加え、複数

の木構造で異なっている部分木を構成する終端記号の多少をも評価配点方法に反映させていることを特徴とする請求項 21 に記載の自然言語処理方法。

【請求項 23】 上記木構造評価工程は、木構造に適用された自然言語パターンでの優先度の情報に加え、所定の優先度に係るノードの位置をも評価配点方法に反映させていることを特徴とする請求項 21 又は 22 に記載の自然言語処理方法。

【請求項 24】 自然言語パターンのユーザ登録工程を有することを特徴とする請求項 14～23 に自然言語処理方法。

【請求項 25】 ユーザ登録の自然言語パターンにシステム登録の自然言語パターン以上の優先度を付与することを特徴とする請求項 24 に記載の自然言語処理方法。

【請求項 26】 構文解析、及び又は、構文生成で得られた木構造が複数の木構造である場合に、複数の木構造で異なっている部分木において、ユーザ登録に係る自然言語パターンを適用された部分木を有する木構造を最優先する木構造評価工程を有することを特徴とする請求項 24 又は 25 に記載の自然言語処理方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は自然言語処理装置及び方法に関し、例えば、翻訳パターンを利用して機械翻訳を行う機械翻訳装置や機械翻訳方法などに適用し得るものである。

【0002】

【従来の技術】従来型の機械翻訳装置において、システムやユーザが登録可能な翻訳知識（システム辞書やユーザ辞書）は、単語辞書、複数単語からなるイディオム、さらに、高性能なものでも「手を焼く」などの動詞と名詞の共起表現など、機械翻訳装置が予め決めている表現に限定されていた。これは、従来型の機械翻訳装置は、原言語の文法と辞書で原言語文を構文解析をし、変換辞書で言語変換し、目的言語の文法と辞書で目的言語文を生成するため、システムの文法に則った形式でしか、システムやユーザは翻訳辞書を登録できないためである。

【0003】さらに、原言語と目的言語は、個別に解析、生成されるため、翻訳知識としてユーザの直感に即した

” [名詞句] を次に示します。”

” Following is [名詞句]. ” (名詞句は変数) (P1)

のような原言語と目的言語の対から構成されるパターンに基づいた辞書も登録できない。

【0004】このような課題を解決したものとして、特開平 5-290082 号公報に記載された翻訳パターンに基づく翻訳方法及び翻訳装置がある。この公報記載の発明は、翻訳知識を文脈自由文法の範囲内で表し、文法規則を原言語のパターンと目的言語のパターンとで対に

したものである。また、文法規則はトライ型の辞書に格納することにより、構文解析の速さを実用レベルにすることを可能にした。さらに、文法規則を原言語のパターンと目的言語のパターンとで対にしたことにより、同期導出が可能となり、言語変換、構文生成は簡単な処理だけで済むようになった。

【0005】これにより、ユーザの直感に即した上述したパターンP1のような原言語と目的言語の対から構成される翻訳パターンに基づいた辞書が登録できるようになり、文脈自由文法の範囲内なら任意の翻訳知識を登録でき、翻訳処理できるようになった。

【0006】

【発明が解決しようとする課題】しかしながら、上記公報では、複数の解（例えば構文解析結果）から最適解を選ぶことについては、最小の構成数（パターン数）の解を選ぶとのみ記述されている。この場合には、構成数がパラメータとなっているので、構成要素の素性や構成要素間の関係を考慮した訳し分けを行うことができない。また、構成数（パターン数）が優先度を規定するものになっているが、単純なパラメータのため、優先度が正しくない恐れがある。

【0007】また、上記公報では、パターンに優先順位を付けても良いと記載されている。しかし、その明確な方式は示されておらず、どのようにパターンに優先順位を付与し、最適解をとらえるかは開示されていない。

【0008】本発明は、以上の点を考慮してなされたものであり、適切な自然言語処理結果を選択し得る自然言語処理装置及び方法を提供しようとしたものである。

【0009】

【課題を解決するための手段】かかる課題を解決するため、第1の本発明は、少なくともパターン名及びパターン構成要素を有する自然言語パターンを利用して、構文解析、及び又は、構文生成を行う自然言語処理装置において、（1）予めパターン辞書に用意されている上記自然言語パターンから、構文解析、及び又は、構文生成時の処理での候補となる1以上の自然言語パターンを抽出する辞書引き手段と、（2）候補の自然言語パターンが木構造に適合しているか否かを検査するパターン検査手段と、（3）適合する場合に、その自然言語パターンを木構造に適用するパターン適用手段とを有することを特徴とする。

【0010】第2の本発明は、少なくともパターン名及びパターン構成要素を有する自然言語パターンを利用して、構文解析、及び又は、構文生成を行う自然言語処理方法において、（1）予めパターン辞書に用意されている上記自然言語パターンから、構文解析、及び又は、構文生成時の処理での候補となる1以上の自然言語パターンを抽出する辞書引き工程と、（2）候補の自然言語パターンが木構造に適合しているか否かを検査するパターン検査工程と、（3）適合する場合に、その自然言語パ

ターンを木構造に適用するパターン適用工程とを有することを特徴とする。

【0011】

【発明の実施の形態】（A）第1の実施形態

以下、本発明による自然言語処理装置及び方法を、機械翻訳装置及び機械翻訳方法に適用した第1の実施形態を図面を参照しながら詳述する。

【0012】（A-1）第1の実施形態の構成

図1は、第1の実施形態の機械翻訳装置の機能的構成を示すブロック図である。なお、実際上は、例えば、パソコンなどの情報処理装置上に、処理プログラムや固定データなどがローディングされて、第1の実施形態の機械翻訳装置が構築される。

【0013】図1において、第1の実施形態の機械翻訳装置は、大きくは、入出力部1、翻訳処理部2及び辞書格納部3から構成される。

【0014】入出力部1は、キーボードやファイル読込装置等の入力部01から翻訳対象となる入力文（原言語文）を受け付ける入力処理部11と、翻訳した結果（目的言語文）をディスプレイやプリンタやファイル格納装置等の出力部02に出力する出力処理部12を有する。

【0015】翻訳処理部2は、入力文の形態素解析を行う形態素解析部21、翻訳パターンを利用して原言語の構文解析を行う構文解析部22、原言語の構文解析の結果に基づいて目的言語による構文の生成を行う構文生成部23、及び、構文生成の結果に基づいて目的言語の形態素生成を行う形態素生成部24からなる。

【0016】構文解析部22は、パターン辞書引きを行う辞書引き部221と、辞書引きされたパターン（原言語パターン）を構築中の木構造に適合するかを検査するパターン検査部223と、その検査結果によってパターンを構築中の木構造に適用するパターン適用部224と、これらの処理部を呼び出し木構造を構築する構文解析器222とを有する。

【0017】辞書格納部3は、原言語の形態素情報が格納されている原言語形態素辞書31、目的言語の形態素情報が格納されている目的言語形態素辞書32、及び、原言語と目的言語のパターンの対からなる翻訳パターンが格納されている翻訳パターン辞書（システムパターン辞書）33からなる。

【0018】（A-2）第1の実施形態の動作

次に、第1の実施形態の機械翻訳装置の翻訳動作（第1の実施形態の機械翻訳方法）を説明する。ここで、図2が、第1の実施形態の機械翻訳装置の翻訳動作を示すフローチャートである。

【0019】なお、第1の実施形態の機械翻訳装置は1文を入力単位としている。また以下の説明においては、"The policeman arrests him."が入力された場合を例に動作を説明する。

【0020】ユーザが、上記入力文を入力部01により

入力すると、入力処理部11が受け取り、翻訳処理部2の形態素解析部21に渡す(ステップS21)。形態素解析部21は、原言語形態素辞書31の情報を用いて、形態素解析処理を行う(ステップS22)。

【0021】形態素解析結果は、図3に示すように、ルートノードを“Node”とした木構造で表現される。複数候補がない形態素の場合には、ルートノードの直下に各形態素の標準形3-1と品詞や変化形などの形態素情報3-2とが付与される。一方、複数候補がある形態素の場合には、orノード3-3の子ノードとして各形態素候補の情報が付与される。例えば、“arrests”は、動詞及び名詞の品詞を取り得るので、複数候補を有する形態素であり、図3に示すように、各形態素候補の情報が付与される。

【0022】このような形態素解析結果が与えられた構文解析部22は、この結果に対し、翻訳パターン辞書33を用いて、構文解析処理を行う(ステップS23)。

【0023】図4は、翻訳パターン辞書33の格納内容(翻訳パターン)の一部を取り出して示したものである。

【0024】図4において、左側が英語パターンを表し、右側が日本語パターンを表し、これら英語パターン及び日本語パターンが対になっている。各言語のパターンは、

〔言語名：パターン名 パターン構成要素〕
からなる。

【0025】言語名は、英語(en)か日本語(ja)かを規定する。

【0026】言語名に続くパターン名は、例えば、VP(動詞句)、NP(名詞句)、N(名詞)等の句構造規則での標識が適用される。

【0027】パターン構成要素は、単語、変数、又は、単語と変数の2以上の並び、からなる。変数は〔任意の数字：パターン名(木構造の下位ノードに対応する)〕で記述される。任意の数字部分は、対となっている原言語及び目的言語パターン間での対応関係を示すものである。構文解析においては、変数に、別のパターンが適用されることにより、パターンは入れ子構造をとることができる(変数が解消される)。また、単語及びパターン名は、符号4-1aや4-1bが付された部分のように、意味情報などの詳細な情報(素性情報)を持つことができる。さらに、単語及びパターン名は、符号4-7aが付された部分のように詳細情報を変数化して、情報の参照をすることもできる。

【0028】このような翻訳パターン(原言語パターン)を利用する第1の実施形態の構文解析処理は、図5のフローチャートに示すように、主として、3つの処理から構成され、ボトムアップに木構造を構築していくものである。

【0029】すなわち、構文解析が終了していないこと

を確認しつつ(ステップS51)、パターン辞書引き処理(ステップS52)、パターン検査処理(ステップS53)及びパターン適用処理(ステップS54)の3つの処理を繰り返す。

【0030】第1の処理であるパターン辞書引き処理(ステップS52)は、形態素解析結果及びそれまでのパターン適用処理の結果から、次に適用の可能性のある翻訳パターンを翻訳パターン辞書33から引き出す処理である。例えば、図3の“him：品詞=prn”(符号3-4)という形態素解析結果より、図4の符号4-4を付した翻訳パターン(原言語パターン)が翻訳パターン辞書33から引かれる。

【0031】第2の処理であるパターン検査処理(ステップS53)は、辞書引きの結果得られた翻訳パターンが現在構築中の木構造に適合するかどうかを、各木構造毎に検査する処理である。

【0032】第3の処理であるパターン適用処理(ステップS54)は、検査の結果、適合すると判定された木構造と翻訳パターンとに基づいて、木構造にそのパターンを実際に適用する処理である。

【0033】図6及び図7を用いて、上述のパターン検査処理及びパターン適用処理を具体的に説明する。

【0034】現在構築中の木構造が図6(a)であり、検査したい翻訳パターンが図6(c)であったとする。変数のパターン名は、NP同士で同じであり、NPがもつ詳細な情報も「意味=人」で一致する。従って、図6(c)に示す翻訳パターンは、図6(a)の木構造に適合するとパターン検査処理で判断され、パターン適用処理によって、その翻訳パターンが適用され、図7に示す木構造が構築される。

【0035】一方、現在構築中の木構造が図6(b)であり、検査したい翻訳パターンが図6(c)であった場合は、変数のパターン名は、NP同士で同じであっても、NPの詳細情報(意味の情報)が異なるので、パターン検査処理で当該翻訳パターンは木構造に適合しないと判断され、当然に、パターン適用処理が実行されることもない。なお、図6(b)において、「意味!=人」は「意味が人ではない」を表している。すなわち、「!」は否定を表している。

【0036】入力文“The policeman arrests him.”に対し、図4に示すような翻訳パターンを適用した構文解析結果は、図8に示すようになる。なお、この入力文に対しては、図4の翻訳パターン4-1、4-3～4-7が適用される。

【0037】このような構文解析結果が構文生成部23に与えられ、構文生成部23によって、構文生成処理(図2のステップS24)が実行される。この構文生成処理では、翻訳パターン辞書33を参照し、原言語(英語)のパターンと対をなしている目的言語(日本語)のパターンを利用して、構文解析結果に対応する日本語の

木構造を得る。図9は、上述した図8の構文解析結果に対応した構文生成結果（木構造）を示している。

【0038】なお、翻訳パターンは、原言語パターンと翻訳パターンとの対になっており、その対応が一意であるので、実際上は、構文解析処理と構文生成処理とがほぼ並行して実行される。また、構文解析結果に対し、パターンを用いない従前からの方法によって構文生成を行っても良い。

【0039】構文生成部23が生成した木構造は形態素生成部24に与えられ、形態素生成部24は形態素生成処理を行う（ステップS25）。形態素生成処理では、構文生成結果を、終端ノードに位置する日本語単語を左から順に並べ、各単語を目的言語形態素辞書32を用いて、動詞の活用形を整えたりなど、各単語の整形を行う。その形態素生成結果（訳文）を出力処理部12が受取り、出力部02によって表示出力したり、ファイルに格納したりする（ステップS26）。

【0040】上述した入力文“The policeman arrests him.”に対しては、「警察官は彼を逮捕する。」という翻訳結果が得られる。

【0041】ここで、上記入力文とは目的語だけが異なる“The policeman arrests the deterioration.”が入力された場合を考える。

【0042】この場合には、目的語にくる名詞の意味が人以外であるので、「警察官は悪化を阻止する。」という翻訳結果が得られる。

【0043】なお、この入力文に対する構文解析では、図4の翻訳パターン4-2、4-3、4-5～4-8が適用される。

【0044】すなわち、上述した図6（a）及び図6（b）に対する構築中の木構造に対し、パターン4-8、4-6が入れ子に挿入された目的語（NP）が人以外を意味するパターン4-2との適合判断により、図6（b）が適合すると判定される。その結果、パターン4-2の情報が盛り込まれた木構造が構文解析で得られ、これにより、arrestの訳としてを阻止するが採用される。

【0045】以上のように、「意味＝人」や「意味！＝人」などのパターン適用条件を、パターンに盛り込んでおくことにより、同じ動詞などを使用した入力文に対して異なる訳語を与えることができる。

【0046】（A-3）第1の実施形態の効果
第1の実施形態の機械翻訳装置及び方法によれば、以下の効果を奏することができる。

【0047】（a）変数による入れ子構造の翻訳パターンを作ることができるだけでなく、変数や単語に意味などのパターン適用条件を与えることによって、条件による訳し分けをすることができる。

【0048】従来の技術の項で言及した公報では、“c

laims \$ and \$”、“claim \$”のように、条件に相当する情報を英語の表現レベルまで落していたが、この実施形態の場合には、英語の表現レベルまで落す必要はない。この実施形態では、“claims [NP：数情報＝複数]”、“claim [NP：数情報＝単数]”となる。

【0049】（b）＝や！＝などパターン適用条件の記述が簡単であり、容易にパターンを作成することができる。例えば、目的語の意味が「人」ならば、訳はA、そうでないならば、訳はBというように、記述形式が人間の直観に即した形式であり、翻訳パターンの作成が容易である。

【0050】（c）“I give these books.”は、構文的には、「私がこれらに本を与える。」と「私がこれらの本を与える。」の2通り存在するが、[en：VP give [l：NP：意味＝人] [2：NP]]という条件をもつパターンを登録することによって、前者の候補はあり得なくなる。このように、意味によるパターン適用条件をパターンに記述することによって、構文の曖昧性を削減することもできる。

【0051】（A-4）第1の実施形態の変形実施形態
上記では、構文解析結果に曖昧性が存在しない例文を用いて説明した。しかし、曖昧性が存在する場合（構文解析結果にorノードが生じる場合（第2の実施形態参照））にも、第1の実施形態を適用できる。例えば、orノードの子ノードのいずれか1つを選んで曖昧性のない構文解析結果を作ってそれを生成したり、又は、曖昧性をもったまま構文生成を行い、複数の翻訳結果を出力したりしても良い。

【0052】第1の実施形態の技術思想は、特定の言語に依存せず、どんな言語間の翻訳にも適用可能である。

【0053】上記では、目的格に位置する名詞の意味による動詞の訳し分けの例を示したが、逆に動詞が持っている意味によって、名詞を訳し分ける場合にも適用できることは勿論であり、また、目的格だけでなく、主格による訳し分けも可能である。さらに修飾する語の意味や、前置詞句内の名詞句の意味による前置詞の訳し分け等、動詞や名詞以外の単語の訳し分けも可能である。さらに、名詞が単数形か複数形かによって名詞の訳語を決めるなど、意味以外のパターン適用条件による訳し分けも可能である。

【0054】パターン検査部などパターンに関する処理は、構文解析処理の外部に設けても良い（第4の実施形態参照）。

【0055】（B）第2の実施形態

次に、本発明による自然言語処理装置及び方法を、機械翻訳装置及び機械翻訳方法に適用した第2の実施形態を図面を参照しながら詳述する。

【0056】（B-1）第2の実施形態の構成

図10は、第2の実施形態の機械翻訳装置の機能的構成

を示すブロック図であり、上述した第 1 の実施形態に係る図 1 との同一、対応部分には同一符号を付して示している。

【0057】第 2 の実施形態の機械翻訳装置は、第 1 の実施形態の構成に加え、構文解析部 22 内にパターン評価部 225 を備えている。

【0058】パターン評価部 225 は、複数候補のパターンが存在する場合に、それらパターンを適用した複数の木構造を評価し、木構造を絞り込むものである。

【0059】なお、翻訳パターン辞書 33 に格納されている翻訳パターンも、第 1 の実施形態のものとは異なっている。この点については、動作の項で明らかにする。

【0060】(B-2) 第 2 の実施形態の動作
この第 2 の実施形態においても、翻訳処理の全体の流れは、図 2 に示す通りであり、原文入力処理、形態素解析処理、構文解析処理、構文生成処理、形態素生成処理及び翻訳結果出力処理の順に実行される。

【0061】ここで、第 2 の実施形態は、第 1 の実施形態に比較して構文解析処理が異なっている。

【0062】図 11 は、第 2 の実施形態の機械翻訳装置での構文解析処理を示すフローチャートであり、第 1 の実施形態に係る図 5 との同一、対応処理ステップには同一符号を付して示している。

【0063】第 2 の実施形態の場合、パターン辞書引き処理 (ステップ S52)、パターン検査処理 (ステップ S53)、パターン適用処理 (ステップ S54) を順次行った後に、構築中の木構造ノードに対応して複数候補のパターンが存在するか否かを判別し (ステップ S55)、複数候補のパターンが存在した場合に、パターン評価処理 (ステップ S56) を行なう。

【0064】新たに追加されたパターン評価処理の具体的な処理内容を説明する前に、図 12 を参照して、この第 2 の実施形態の翻訳パターン辞書 33 に格納されている翻訳パターンを説明する。

【0065】第 2 の実施形態の翻訳パターンは、先頭の " [" と末尾の "] " と挟まれている第 1 の実施形態と同様なパターンに加え、パターンの先頭に " ! " が付与されたパターン 12-9 やパターンの先頭に " * " が付与されたパターン 12-12 がある。" ! " や " * " は、パターンの優先度を表すための特別な記号である。

【0066】図 13 は、パターンの優先度を規定するファイルの例であり、翻訳パターン辞書 33 又はパターン評価部 22 が格納しているものである。ここでは、ファイルの先頭から順に優先度が高いことを示している。また、□ は記号が何もない場合の順位である。従って、この例での優先順位は " ! " が付与されているパターン、優先度の記号なしのパターン、" * " が付与されているパターンとなる。

【0067】図 12 において、定冠詞の後に、名詞及び

る英語形態素列に関しては、2 番目の単語を形容詞としたときには、パターン 12-9 が適用され、2 番目の単語を名詞としたときには、パターン 12-10 が適用される。図 12 に示すように、パターン 12-9 がパターン 12-10 より優先度が上げられているので、動作の詳細は後述するが、両パターンが適用可能な場合であれば、パターン 12-9 が優先される。

【0068】また、図 12 において、パターン 12-11 及びパターン 12-12 は共に文 (S) に関するパターンであるが、意味条件が付されていない汎用的な方であるパターン 12-12 の優先度が低くなっている。動作の詳細は後述するが、意味条件が付されたパターン 12-11 が適用されない場合に、汎用的なパターン 12-12 が採用されることになる。

【0069】パターン評価処理 (ステップ S56) では、このようなパターンの優先度を利用しながら複数候補のパターンが適用された複数の木構造の評価を行う。

【0070】図 14 は、パターン評価処理の詳細を示すフローチャートである。パターン評価部 224 は、まず、ルートノードの詳細情報 (パターン名の詳細情報) が一致している複数の木構造候補があるか、ある場合には、ルートノードに高い優先度が付いた木構造候補があるか否かを判別する (ステップ S141)。ここで、高い優先度とは候補の中で相対的に高い優先度を言い、必ずしも " ! " が付与されたものに限定されない。そして、高い優先度が付いた木構造候補あれば、その木構造候補以外を候補から外す (ステップ S142)。また、パターン評価部 224 は、子ノードの構成要素が一致している複数の木構造候補があるか、ある場合には、子ノードに低い優先度が付与されている木構造があるか否かを判別する (ステップ S143)。ここで、低い優先度とは候補の中で相対的に低い優先度を言い、必ずしも " * " が付与されたものに限定されない。そして、低い優先度が付与されている木構造があれば、その木構造を候補から外す (ステップ S144)。

【0071】次に、第 2 の実施形態の構文解析処理、特に、パターン評価処理について、入力文 " He arrests the present deterioration. " を例に説明する。この入力文に対する解析では、上述した図 12 に示した翻訳パターンが適宜利用される。

【0072】図 15 (a) 及び (b) はそれぞれ、入力文中の " the present deterioration " の部分に対し、図 12 に示すパターン 12-9 及び 12-10 が適用されてできた候補 (評価対象) としての木構造である。図 15 (a) に示す木構造には、翻訳パターン 12-9 に付与されている " ! " の記号が付いている。パターン適用処理の終了後、図 15 (a) 及び (b) に示すような木構造のルートノードが等しく、かつ木構造を構成している単語が等しい木構造が存

在すれば、これらの木構造がパターン評価処理に渡される。

【0073】パターン評価処理では、図13に示した優先度記号ファイルの内容を予め読み込んでおき、それに従って、パターンの適用を評価する。仮に、記号なしより高く評価したい記号がついた木構造が存在すれば、ルートノードがもつ詳細情報が等しいときのみ、最も高い記号をもつパターンのみを採用し、それ以外は構築途中の木構造の候補から削除する。

【0074】図15の例の場合には、図15(a)の木構造と図15(b)の木構造の詳細情報が等しいので、記号がついていない図15(b)の木構造は、優先度を高める記号が付いた図15(a)の木構造より優先度が低いので、木構造の候補から削除される。

【0075】一方、図16は、記号なしのパターンより優先度を下げる記号“*”が付いている木構造を複数候補の中にもつ例である。優先度を下げる記号の評価の場合には、木構造の子ノードの構成要素が等しいことが条件となる。図16(a)に示すものはパターン12-11が適用された記号なし木構造であり、図16(b)に示すものはパターン12-12が適用された優先度を下げる記号“*”が付いた木構造であり、これら木構造は、子ノードの構成要素がNP、VPと等しく、かつ、図16(b)に示す木構造は“*”が付いているので、この木構造が候補から外される。

【0076】図17は、入力文“*He arrests the present deterioration.*”に対する最終的な構文解析結果を示している。

【0077】この構文解析結果に対応する日本語木構造が構文生成処理で生成され、形態素生成処理を経て、最終的に、「彼は現在の悪化を阻止する。」の翻訳結果が出力される。

【0078】(B-3)第2の実施形態の効果
第2の実施形態の機械翻訳装置及び方法によれば、第1の実施形態と同様な効果を奏すると共に、さらに、以下のような効果をも奏する。

【0079】(a)構文的には正しい複数の木構造候補が存在する場合、その中のある木構造を優先的に適用させて翻訳したい場合がある。この第2の実施形態では、優先度を上げる記号“!”を優先的に適用させたいパターンに付与することで、これが実現できる。

【0080】(b)木構造解析の途中で、パターンの条件は等しく、その優先度に差がある複数候補のパターンが存在した場合に、優先度が低いパターンをその候補から外すことにより、解析候補が減り、構文解析や構文生成にかかる処理時間を削減することができる。

【0081】(c)優先度下げ記号が付いた意味条件がついていないパターンを登録することによって、意味条件が適用する場合には、意味条件がりのパターンのみが適用され、意味条件が適用しない場合にだけ、優先度

下げ記号がついた意味条件なしのパターンに適用するようになる。これにより、常に意味条件なしのパターンを適用するという解析の無駄をなくことができ、解析にかかる処理時間が削減される。さらに、意味条件を適用しない場合も、構文解析の失敗にならず、尤もらしい(汎用的な)訳語を得ることができる。

【0082】(d)尤もらしい(汎用的な)訳語に対応したパターンに優先度を上げる記号を付与することで、意味条件がないパターンに適用する場合には、厳しい条件をもつ特殊な訳語ではなく、汎用的な訳語で翻訳することができる。

【0083】(B-4)第2の実施形態の変形実施形態
優先度を示す記号はどんな種類の記号(単なる記号だけでなく、数字や括弧の種類や文字列などを含む)であっても構わない。また、記号ではなく、他の形式の情報であっても良い。

【0084】優先度を上げる記号として、!と+など複数設け、!を優先させるというように、優先度を順に複数用意することもできる。

【0085】(C)第3の実施形態

次に、本発明による自然言語処理装置及び方法を、機械翻訳装置及び機械翻訳方法に適用した第3の実施形態を図面を参照しながら詳述する。

【0086】(C-1)第3の実施形態の構成

図18は、第3の実施形態の機械翻訳装置の機能的構成を示すブロック図であり、上述した第2の実施形態に係る図10との同一、対応部分には同一符号を付して示している。

【0087】第3の実施形態の機械翻訳装置は、第2の実施形態の構成に加え、木構造評価部25を備えている。

【0088】木構造評価部25は、構文解析により完成した原言語の構文解析結果として複数の候補が存在した場合に、どの候補が最も適切かをパターンがもつ情報などによって評価(点数計算)して、最適な候補(木構造)を選択して構文生成部23に与えるものである。又は、複数の候補を評価値順に並べ替えて構文生成部23に与えるものである。

【0089】(C-2)第3の実施形態の動作

図19は、第3の実施形態の機械翻訳装置の翻訳動作を説明するフローチャートであり、第1の実施形態に係る図5との同一、対応処理ステップには同一符号を付して示している。

【0090】この第3の実施形態においては、原文入力処理(ステップS21)、形態素解析処理(ステップS22)、構文解析処理(ステップS23)、木構造評価処理(ステップS27)、構文生成処理(ステップS24)、形態素生成処理(ステップS25)及び翻訳結果出力処理(ステップS26)がこの順に実行される。すなわち、第1や第2の実施形態に比較して、木構造評価

処理（ステップS27）が増えている。

【0091】図20は、木構造評価処理を示すフローチャートである。木構造評価部25は、構文解析結果を受け取ると、これをボトムアップに処理して評価点を計算する。木構造評価部25は、枝の先端側から非終端ノードを取り出しては（ステップS2001）、それがorノードかを判別し（ステップS2002）、orノードでなければ現ノードの点数を計算し（ステップS2005）、orノードであれば、子ノードの点数のうち最高点をノードの点数とした後（ステップS2003）、点数順にノードを並び換える（ステップS2004）。このような処理を、点数計算が終了していない非終端ノードがなくなるまで繰り返す。なお、具体的な点数の計算方法は後述する。

【0092】以下、木構造評価処理を、入力文“*It arrests an attention.*”を例に具体的に説明する。

【0093】図21は、この入力文に対する構文解析結果である。この図21は、orノードが2個存在するので、4個の構文解析結果候補（4個の木構造をまとめて示している）を含んでいる。図21において、点線で引き出された数字は木構造評価処理において計算された値であり、構文解析結果では含まれていない。なお、構文解析処理で用いられた翻訳パターン辞書33は第2の実施形態のもの（図12）に準じているとする。但し、優先度上げ記号として“!”と“+”との2種類があり、前者が2段階の優先度上げを表し、後者が1段階の優先度上げを表しているとする。

【0094】図22は、評価点の計算方法の定義ファイルを示している。評価点計算方法定義ファイルは、例えば、木構造評価部25に格納されている。

【0095】評価点計算方法定義ファイルは、評価項目名フィールド221と、評価部分フィールド222と、評価点フィールド223とからなる。

【0096】評価項目名フィールド221には、パターンに係る優先度の上げ下げの記号“!”や“+”や“*”や“node”（一般的な優先度ノード）だけでなく、「木構造の終端記号の数」を意味するterminalや、「ユーザ登録に係るもの」を意味するuser（例えば形態素辞書やパターン辞書に対する登録）などの評価項目名が記述されている。

【0097】評価部分フィールド222には、自ノード（例えばorノード）より下の枝の全てが評価部分であることを表すtotalや、自ノード（例えばorノード）の直下の子ノードが評価部分であることを表すchildなどの評価部分が記述されている。

【0098】評価点フィールド223には、評価点が記述されている。ここで、評価部分がtotalであるものに対しては、該当する評価項目（例えば記号）の数（Nとする）に応じた計算式（例えば、 $1000-10$

ON）で評価点が記述されている。なお、 $1000-100N$ は、数学的には、 $1000-100\times N$ を意味している。評価部分がchildであるものに対しては、固定の評価点が記述されている。

【0099】図22に示す評価点計算方法定義ファイルは、ユーザ登録に係る事項を最も評価し、次に、終端記号数を評価し、その後は、パターンの優先度の順位を評価するような構成となっている。なお、最高優先度記号“!”については、orノード直下でのノードに付されているか否かをも大きな評価ウェイトとしている。

【0100】上述した図20に示した木構造評価処理において、ステップS2005の点数計算は、評価点計算方法定義ファイルの規定内容に従ってなされる。

【0101】図21により具体的な評価点の計算方法を説明する。

【0102】ノード21-1の+Nは、“+”記号が付与されており、また、非終端記号数がattentionだけの1個であるので、図22の定義ファイルにおける225及び230のレコードが適用されて評価点は90000001点となり、一方、ノード21-2は、“+”記号が付与されておらず、また、非終端記号数がattentionだけの1個であるので、図22の定義ファイルにおける225のレコードが適用されて評価点は90000000点となる。その結果、次のorノード21-3では、高い方のノード21-1の評価点が採用される。

【0103】また、ノード21-4のVPは“arrests an attention”が一塊りであって終端記号数（終端ノード数）も子ノード数の1個であるので、ファイルのレコード225及び228が適用されて評価点は90000900点となり、一方、ノード21-5のVPは、終端記号数（終端ノード数）が3個、子ノード数が4個、“+”記号が枝の先端側に1個あるので、ファイルのレコード225、228及び2210が適用されて、評価点は7000601点となる。その結果、次のorノード21-6ではノード21-4の点数が採用される。

【0104】以上のようにして、評価点の計算が終了すると、ノードの順番を点数順に並び変える。このようにして、並べ替えた後の木構造を図23に示す。

【0105】構文生成処理は、この木構造に対応する日本語（目的言語）での木構造を生成する。仮に、英語の木構造にorノードがあれば、orノードの子ノードで最も先頭に位置する英語ノードに対応する日本語ノードのみが選ばれて、生成処理が行われる。その後、形態素生成処理が行われ、その結果、「それは興味を引く。」という翻訳結果が出力される。

【0106】なお、仮に、ユーザが複数の翻訳結果候補を出力して欲しい場合には、orノードの子ノード全てに対応する日本語木構造を利用して構文生成、形態素生

成処理を行う。この場合の翻訳結果は、「それは（興味を引く|注意を阻止する）。」となる。この場合、前者の方がシステムは評価していることを示している。

【0107】(C-3)第3の実施形態の効果
第3の実施形態の機械翻訳装置及び方法によれば、第1及び第2の実施形態と同様な効果を奏すると共に、これに加え、以下の効果を奏することができる。

【0108】により以下の効果を有する。

【0109】(a) 木構造全体で、パターンに付与されている種々の優先度記号や、パターンの詳細情報を用いて評価することにより、よりきめ細かな訳し分けや構文解析との絡みによる訳し分けが可能となる。

【0110】(b) この第3の実施形態によれば、種々の優先度記号やパターン詳細情報の優先順位を決めておくことにより、優先順位に従った翻訳結果を期待でき、翻訳結果も予測しやすい。

【0111】(c) 終端パターンの数など、優先度記号以外の条件も、同じ計算式で同時に評価することができる。

【0112】(d) 例えば、“...today...”の解析結果において、名詞の候補しか存在しない場合は名詞として翻訳するが、名詞と副詞の両方の候補が存在した場合には、副詞として翻訳したい、というような場合、優先したい品詞（この場合副詞）のパターンに優先度を上げる記号を付与することで、多品詞の優先順位を制御することができる。

【0113】(C-4)第3の実施形態の変形実施形態
第3の実施形態で用いる評価点の計算方法を制御する優先度記号は、第2の実施形態の優先度記号を併用してもかまわなく、また、異なるようにしても良い。前者の場合において、記号の付与によってパターン評価処理で候補を削減できるものは削減し、それ以外を点数計算で優先度を決めるといふ、1つの記号で二段階の優先度処理が行われることになる。上記説明は、このような場合を意識している。

【0114】評価点の計算方法を規定する記号は、いくつ付与してもどんな種類であっても構わない。また、複数種類の記号を同時に付与することも可能である。

【0115】点数の計算式及び各評価項目の配点は、どんな式でもどんな配点でも構わない。また、簡単に変更し得るようにしても良い。木構造全体の記号の数の優先順位より、orノード直下のノードに付与された記号の優先順位を高くするなど、木構造のレベルを考慮に入れた点数計算をすることも可能である。

【0116】終端ノードの数だけでなく、他の木構造やパターンの特徴に従って優先度を定めることができる。例えば、第1の実施形態の目的語の名詞句の意味の数が少ないものほど優先するように、記号を付与すると、複数のパターンが条件に適用した場合でも、常に、条件の厳しいパターンが優先されて翻訳されるというように、

翻訳結果の出力に一定の原則を与えることができる。

【0117】(D)第4の実施形態

次に、本発明による自然言語処理装置及び方法を、機械翻訳装置及び機械翻訳方法に適用した第4の実施形態を図面を参照しながら詳述する。

【0118】(D-1)第4の実施形態の構成

図24は、第4の実施形態の機械翻訳装置の機能的構成を示すブロック図であり、第3の実施形態に係る図18との同一、対応部分には同一符号を付して示している。

【0119】第4の実施形態の機械翻訳装置は、パターン検査部251、パターン適用部252及びパターン評価部253の部分を構文解析部22の外部に独立したパターン処理部26として備え、構文生成部23内の構文生成器231が利用し得るようにしている。

【0120】すなわち、パターンの検査、適用、評価を構文解析時だけでなく、構文生成時に行うことができるようにしている。

【0121】なお、翻訳パターンとして、同一の原言語パターンに対応する目的言語パターンが複数のこともある。また、図4では示していないが、目的言語パターンにもパターン適用条件などを当然に付与することができる。この第4の実施形態は、このようなことを前提としている。

【0122】(D-2)第4の実施形態の動作

第4の実施形態が既述した実施形態と異なる動作は、構文生成処理である。そこで、以下では、図25のフローチャートを参照しながら、第4の実施形態の構文生成処理での特徴的な処理を説明する。

【0123】従来の方式では、英語（原言語）の木構造の構築とほぼ並行して日本語（目的言語）の木構造も構築されるため（単純なパターンを利用した木構造の置き換え）、特別な構文生成処理は行われない。そのため、日本語の木構造においてパターン適用条件のずれが生じてもそれを検知することができない。

【0124】この第4の実施形態は、英語の木構造の構築とほぼ並行して構築される日本語の木構造において、複数候補ノードが発生した場合に、パターン検査処理、パターン適用処理、パターン評価処理を行うことによって、複数候補を解消しようとしたものである。英語の木構造に対して、例えば、日本語の木構造を従来の構文生成処理（パターンを用いた単純な置き換え）で得た仮りの日本語（目的言語）の木構造に対して、図25が実行される。

【0125】まず、構文生成器231は、ルートノードからトップダウンに仮の木構造から非終端記号を1個取り出し（ステップS2501）、それがorノードか否かを判別する（ステップS2502）。

【0126】orノードがあれば、orノードの子ノードのパターンと自ノードを含めた親の木構造をパターン検査部223に渡し、パターン検査処理及びパターン適

用処理を実行させる（ステップ S 2503）。パターン検査処理が終了したときには、その検査結果に応じて、パターン適用部 224 によるパターン適用処理が実行される（ステップ S 2504）。その後、パターン適用処理により or ノードが解消されたか否かを判別し（ステップ S 2505）、解消されていないならば、パターン評価部 225 にパターン評価処理を実行させる（ステップ S 2507）。

【0127】取り出した非終端記号が or ノードでない場合には、そのノードをそのまま適用させる（ステップ S 2506）。

【0128】このような処理を非終端記号が取り出せなくなるまで繰り返し、日本語（目的言語）の木構造を完成（生成）させる。

【0129】なお、パターン検査処理、パターン適用処理及びパターン評価処理は、既述した実施形態とほぼ同様な処理である。

【0130】以下、入力文 “the intellectual woman” を例に、第 4 の実施形態での構文生成処理を説明する。

【0131】図 26（a）は、この入力文に対する構文解析結果（英語での木構造）を示し、図 26（b）は、それに対応する日本語での木構造を示している。図 26（b）の日本語木構造は英語木構造の対応結果から作られた仮の木構造である。この仮の木構造に対して、上述した図 25 に示す処理が実行される。

【0132】図 26（b）に示す木構造では or ノードが含まれており、パターン検査処理やパターン適用処理が行われる。

【0133】ここで、図 27（b1）に示すような翻訳パターンが記述されている場合であって、パターン検査処理及びパターン適用処理に供する親の木構造が図 27（a）に示すような場合であれば、パターン検査でパターンは適合するという結果が得られ、図 27（b1）の日本語パターンが適用される。その結果、or ノードは解消されて次のノードに進む。

【0134】しかし、図 27（b2）に示すような翻訳パターンも、すなわち、日本語パターンが英語パターンに対応して 2 種類記述されている場合であって、両パターンが適用できる場合であれば、パターン評価処理が実行される。

【0135】2 種類の日本語パターンには優先度記号が付与されているので、ここで、1 個に絞り込まれる。又は、優先順位が付与される。

【0136】図 28 は、上記入力文に対して優先順位が付与した場合の構文生成処理結果を示している。この結果が形態素生成部に渡されて、「理知的な女性」又は「理知的な女性 | 知力を要する女性」の翻訳結果が得られる。

【0137】（D-3）第 4 の実施形態の効果

第 4 の実施形態の機械翻訳装置及び方法によれば、既述した実施形態の効果に加えて、以下の効果を奏することができる。

【0138】（a）原言語に関する情報だけでなく、目的言語の情報によっても訳し分けをすることができる。

【0139】（b）上記翻訳パターン辞書による双方向の翻訳を実施した場合、原言語、目的言語パターンの両方に優先度を付与することによって、それぞれの優先度を解析側、生成側で再び利用することができる。

【0140】（D-4）第 4 の実施形態の変形実施形態上記では、パターン生成時に複数候補を減らすという、第 2 の実施形態に類似の方式を説明したが、構文解析器は全ての候補を求め、その後、構文生成木の点数計算をするという第 3 の実施形態に類似した方式をとっても良い。また、両方の方式をとっても良い。

【0141】原言語から目的言語の翻訳の逆（つまり目的言語から原言語への翻訳）を考えた場合、目的言語のパターンに付与した優先度は解析時に利用できる。その場合、原言語側のパターンに付与した優先度は、生成時に利用できる。

【0142】（E）第 5 の実施形態

次に、本発明による自然言語処理装置及び方法を、機械翻訳装置及び機械翻訳方法に適用した第 5 の実施形態を図面を参照しながら詳述する。

【0143】（E-1）第 5 の実施形態の構成

図 29 は、第 4 の実施形態の機械翻訳装置の機能的構成を示すブロック図であり、第 4 の実施形態に係る図 24 との同一、対応部分には同一符号を付して示している。

【0144】この第 5 の実施形態は、入出力部 1 に、ユーザの入力する翻訳パターンを受け付けるユーザ辞書登録処理部 13 を備え、また、辞書格納部 3 にその翻訳パターンを格納するユーザパターン辞書 34 を備えている。すなわち、ユーザ登録（削除）機能を搭載したものである。

【0145】そのため、構文解析部 22 などは、システムパターン辞書 33 とユーザパターン辞書 34 の両方を参照することになる。

【0146】（E-2）第 5 の実施形態の動作

ユーザ登録に係るものが翻訳パターンであるため、条件等の詳細な情報の直接的な登録は専門知識が必要とされるが、ユーザインターフェース等を備えることにより、単語（形態素）その他の要素の登録と同様な処理により登録することができる。但し、以下のような点が異なっている。

【0147】”Following is [名詞句]” を “[名詞句] を次に示します。” をユーザが登録する場合について説明する。

【0148】この場合の翻訳パターンは、

[en:S [following is [1:N P].]

[j a : S [1 : NP] を次に示します。] となる。
 【0149】仮に、名詞句に位置する単語が人以外の場合のみこのパターンを適用したい場合には、[e n : S [following is [1 : NP : 意味 ! = 人] .] [j a : S [1 : NP] を次に示します。] というように意味条件も記述して登録する。この場合において、ユーザによる任意の入力を待ち受けるだけでなく、ユーザ辞書登録処理部13が、例えば、NPなどの場合に条件設定を行うか否かを問うメッセージや、条件の設定例などを表示して、条件を取り込むようにしても良い。

【0150】また、ユーザ登録に係るパターンには、最高の優先度上げ記号を常に付加して登録することにする。

【0151】さらに、上述したように、図22の評価点計算方法定義ファイル(第3の実施形態に係る)では、user という評価項目名のレコード221に最も高い評価点を付与するようにしており、木構造の評価時にユーザパターン辞書のパターンに係るものが最優先され、その翻訳結果が得られる。

【0152】ここで、user はユーザパターンを示し、child は子ノードの直下を示す。レコード22-1は、or ノード直下のユーザパターンに対し、最高点(100000000点)を与える、を意味する。これにより、ユーザパターンは、全ての候補に対して、優先度が高くなり、ユーザが登録したパターンは、必ず翻訳結果に反映されるようになる。

【0153】(E-3) 第5の実施形態の効果
 第5の実施形態の機械翻訳装置及び方法によれば、既述した実施形態の効果に加えて、以下の効果を奏することができる。

【0154】(a) ユーザが作成する翻訳パターン辞書にも、変数に意味やその他の条件を付与することができる。これにより、一般化したパターンによる登録ができ、登録するパターン数は少なく済む。

【0155】(b) ユーザパターン毎に優先度記号を付与することも可能であり、ユーザによる訳語や訳文の制御も可能である。

【0156】(c) ユーザパターンに係る評価点も他のパターンに対する計算と同じ点数計算で行うため、ユーザパターン辞書の優先度(優先度記号)などの変更にも容易に対応することができる。

【0157】(d) ユーザは、候補訳語を全て出力することによって、ユーザパターンを用いない翻訳結果もユーザパターンを用いた翻訳と同時に得ることができる。

【0158】(E-4) 第5の実施形態の変形実施形態
 第5の実施形態では、ユーザパターン辞書を最優先にしたが、最少ノード数を優先し、次に、ユーザパターンを優先する等、優先順位は、適宜変更することができる。

【0159】複数のユーザパターン辞書を作成して、そ

れぞれの辞書に優先順位を付けることができる。

【0160】ユーザパターン辞書においても優先度を上げたり、下げたりする記号を付与することができる。

(その場合は、ユーザパターン辞書内における優先順位付けとなる。)

ユーザパターン辞書入力処理部において、ユーザは、[e n : ...] [j a : ...] などのような翻訳パターンを直接入力するのではなく、パターン名、単語と変数を簡単に入力するユーザインターフェースを設けることにより、ユーザパターン辞書の作成を容易にすることができる。

【0161】第5の実施形態では、ユーザパターンの評価を木構造評価部で行うことによってユーザパターン辞書を優先して翻訳する方式を示したが、パターン評価部において、ユーザパターンとそれ以外の候補が存在した場合、ユーザパターンを適用して、それ以外の候補は適用しないという方式でも実現可能である(第2の実施形態参照)

(F) 他の実施形態

本発明の技術思想の適用対象は、機械翻訳装置や方法に限定されず、自然言語パターンを利用して構文解析を行う自然言語処理装置及び方法や、自然言語パターンを利用して構文生成を行う自然言語処理装置及び方法に適用することができる。

【0162】この場合において、構文解析を自然言語パターンを利用して行い、構文生成は自然言語パターンを利用しない方法で行うものであっても良く、その逆であっても良い。また、構文解析のみを必要とし、構文生成を実行しない装置や、その逆の装置などにも本発明を適用することができる。

【0163】

【発明の効果】以上のように、本発明によれば、適切な自然言語処理結果を提供し得る自然言語処理装置及び方法を実現できる。

【図面の簡単な説明】

【図1】第1の実施形態の機械翻訳装置の機能的構成を示すブロック図である。

【図2】第1の実施形態の機械翻訳装置の翻訳動作を示すフローチャートである。

【図3】第1の実施形態の形態素解析結果例を示す説明図である。

【図4】第1の実施形態の翻訳パターン辞書33の格納内容例を示す説明図である。

【図5】第1の実施形態の構文解析処理の詳細を示すフローチャートである。

【図6】第1の実施形態の構文解析処理のパターン検査処理及びパターン適用処理の具体例の説明図である。

【図7】図6のパターン適用処理後の木構造を示す説明図である。

【図8】第1の実施形態の入力文例に対する構文解析結

果を示す説明図である。

【図 9】図 8 の構文解析結果に対する構文生成結果を示す説明図である。

【図 10】第 2 の実施形態の機械翻訳装置の機能的構成を示すブロック図である。

【図 11】第 2 の実施形態の構文解析処理の詳細を示すフローチャートである。

【図 12】第 2 の実施形態の翻訳パターン辞書 33 の格納内容例を示す説明図である。

【図 13】第 2 の実施形態のパターンの優先度記号ファイルを示す説明図である。

【図 14】第 2 の実施形態のパターン評価処理の詳細を示すフローチャートである。

【図 15】第 2 の実施形態のパターン評価処理の対象となる木構造候補を示す説明図 (1) である。

【図 16】第 2 の実施形態のパターン評価処理の対象となる木構造候補を示す説明図 (2) である。

【図 17】第 2 の実施形態の構文解析結果の例を示す説明図である。

【図 18】第 3 の実施形態の機械翻訳装置の機能的構成を示すブロック図である。

【図 19】第 3 の実施形態の機械翻訳装置の翻訳動作を示すフローチャートである。

【図 20】第 3 の実施形態の木構造評価処理を示すフローチャートである。

【図 21】第 3 の実施形態の木構造評価処理に供する構文解析結果例を示す説明図である。

【図 22】第 3 の実施形態の木構造の評価点計算方法定義ファイルを示す説明図である。

【図 23】第 3 の実施形態の木構造評価処理語の構文解

析結果を示す説明図である。

【図 24】第 4 の実施形態の機械翻訳装置の機能的構成を示すブロック図である。

【図 25】第 4 の実施形態の構文生成処理の特徴処理を示すフローチャートである。

【図 26】第 4 の実施形態の構文生成処理の特徴処理を行う前の木構造を示す説明図である。

【図 27】第 4 の実施形態のパターン検査処理及びパターン適用処理の説明図である。

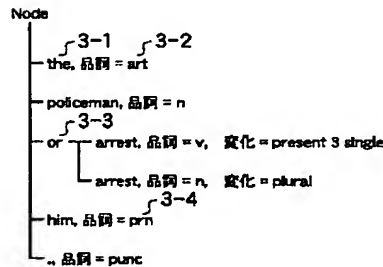
【図 28】第 4 の実施形態の構文生成結果例を示す説明図である。

【図 29】第 5 の実施形態の機械翻訳装置の機能的構成を示すブロック図である。

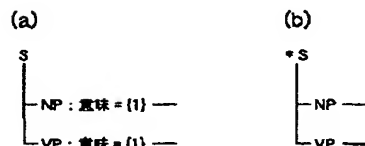
【符号の説明】

- 1…入出力部、
- 2…翻訳処理部、
- 3…辞書格納部、
- 13…ユーザ辞書登録処理部、
- 21…形態素解析部、
- 22…構文解析部、
- 23…構文生成部、
- 24…形態素生成部、
- 25…木構造評価部、
- 26…パターン評価部、
- 221…辞書引き部、
- 223、261…パターン検査部、
- 224、262…パターン適用部、
- 225、263…パターン評価部、
- 33…翻訳パターン辞書 (システムパターン辞書)、
- 34…ユーザパターン辞書。

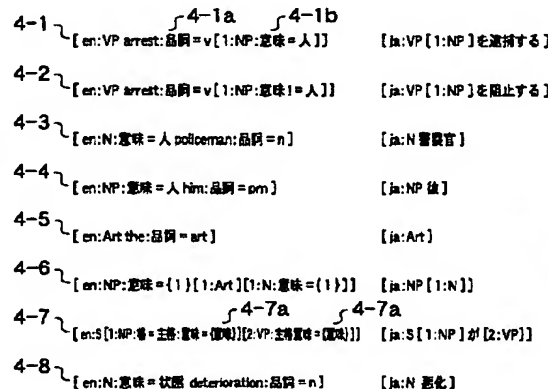
【図 3】



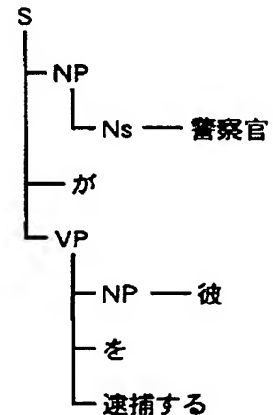
【図 16】



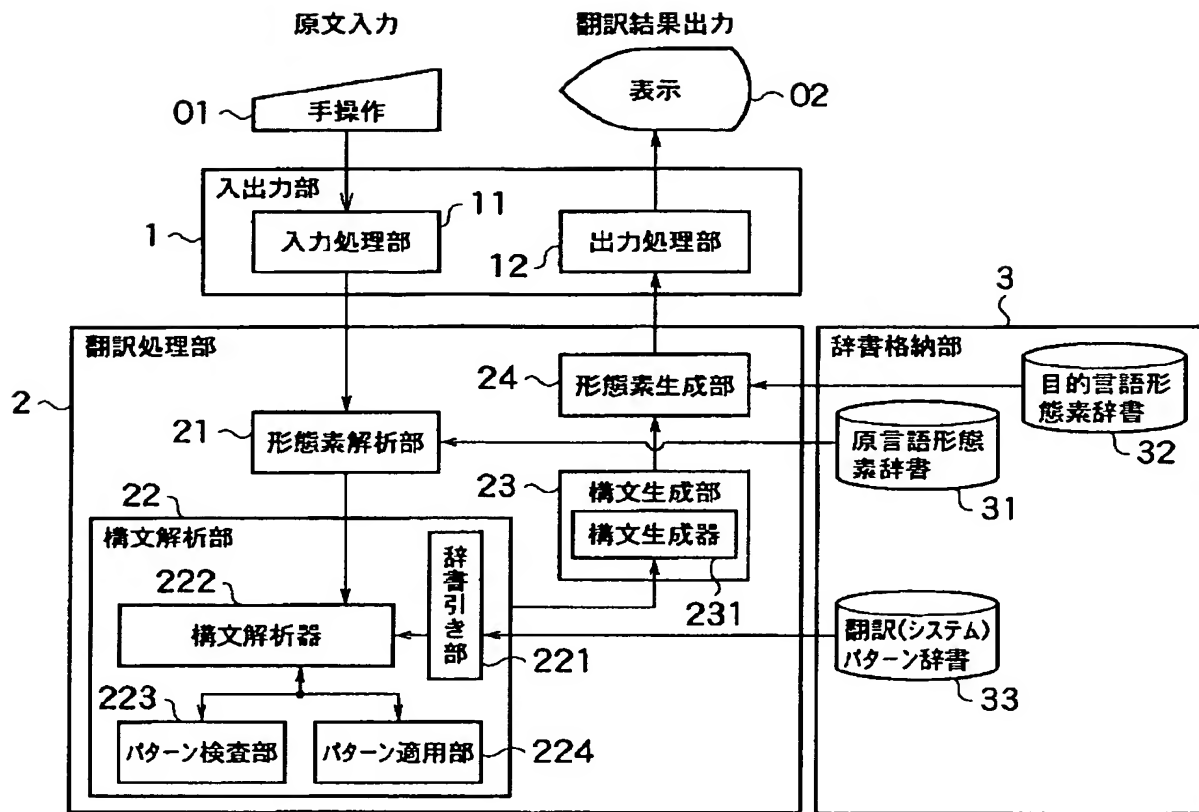
【図 4】



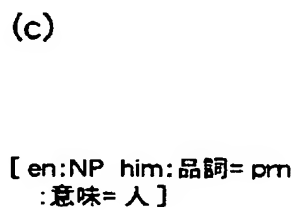
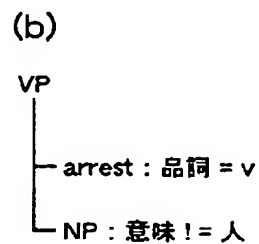
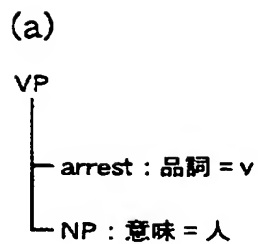
【図 9】



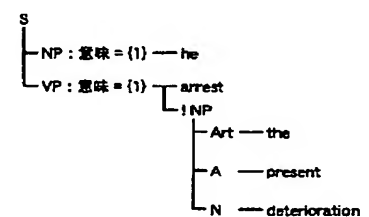
【図1】



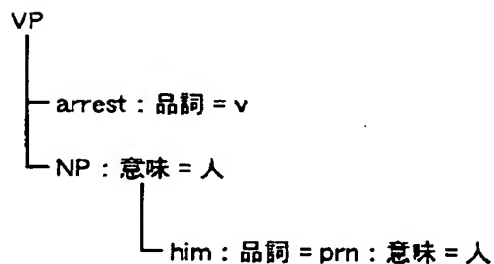
【図6】



【図17】



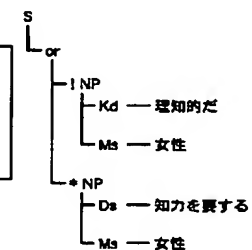
【図7】



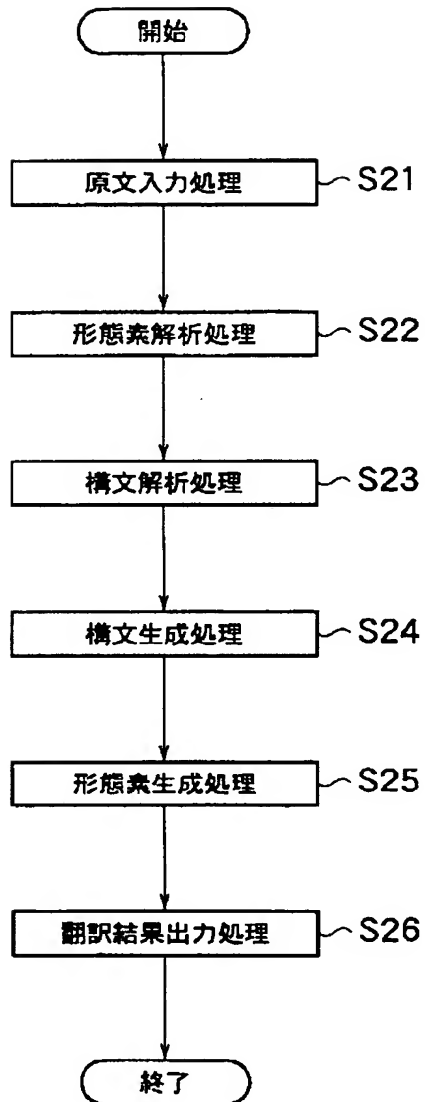
【図13】



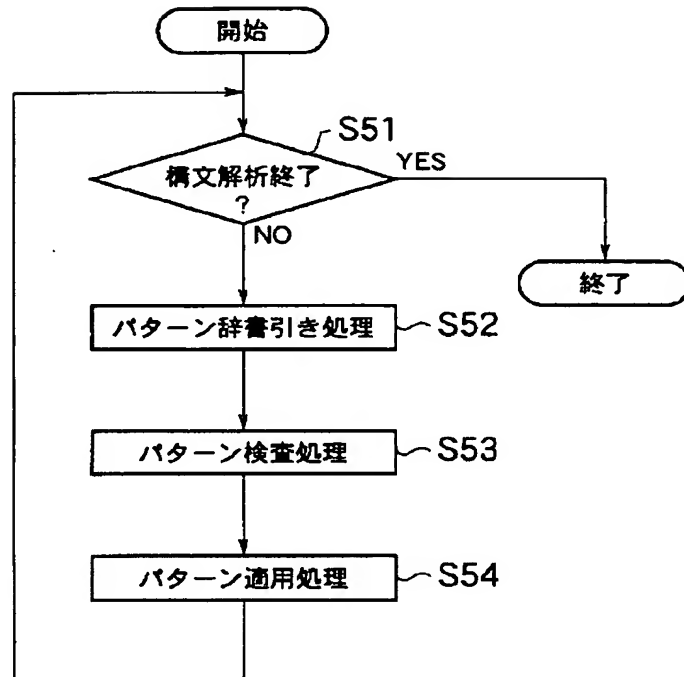
【図28】



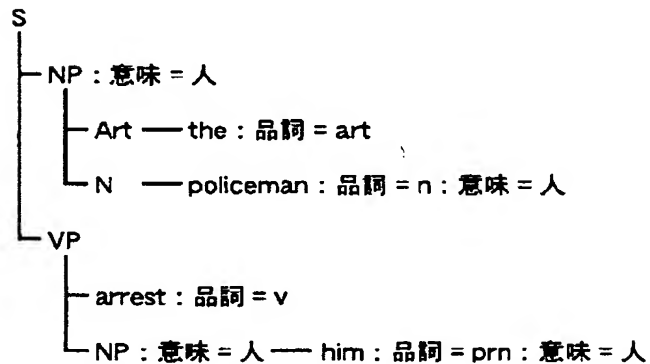
【図2】



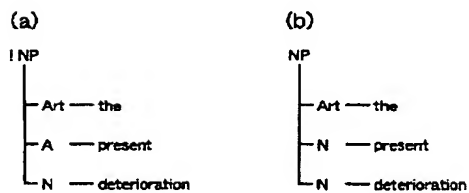
【図5】



【図8】



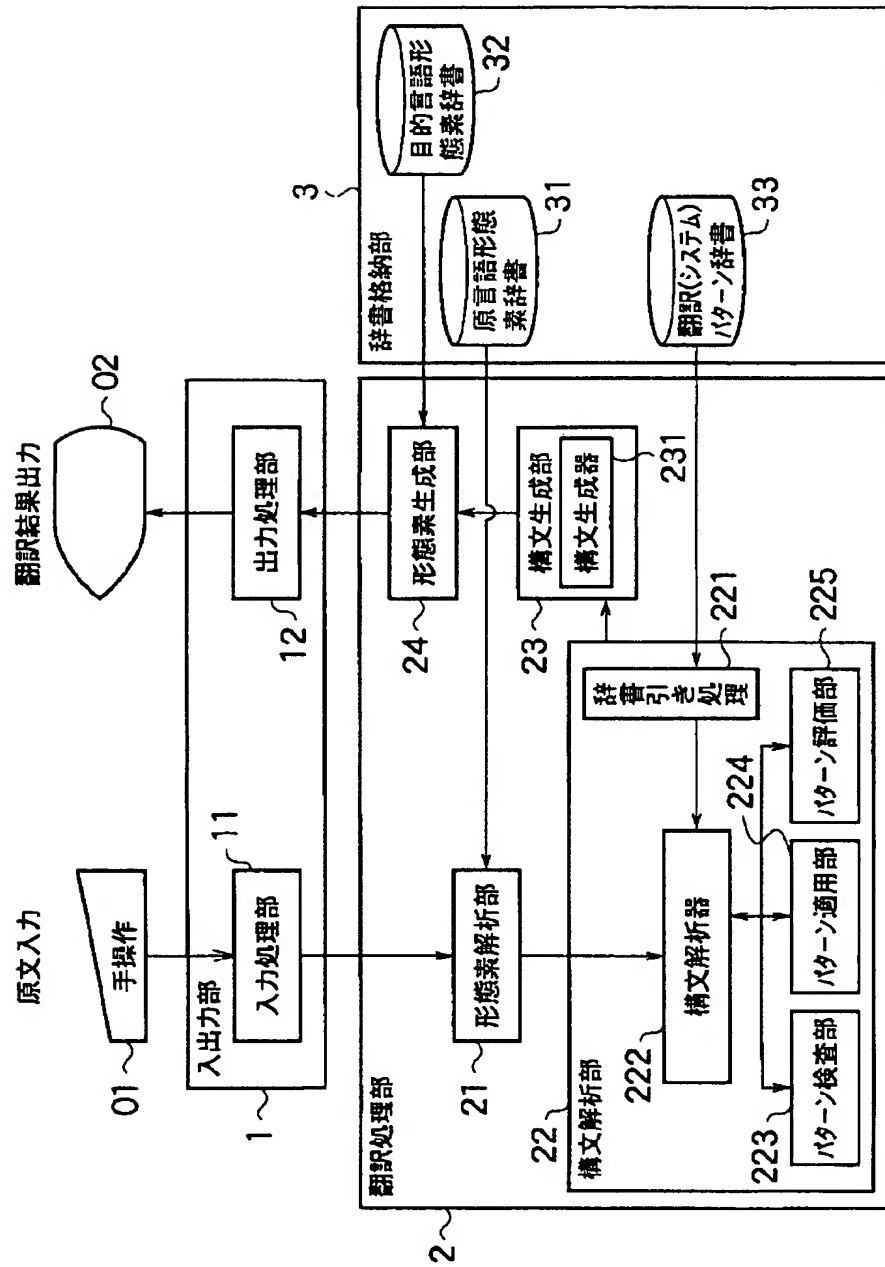
【図15】



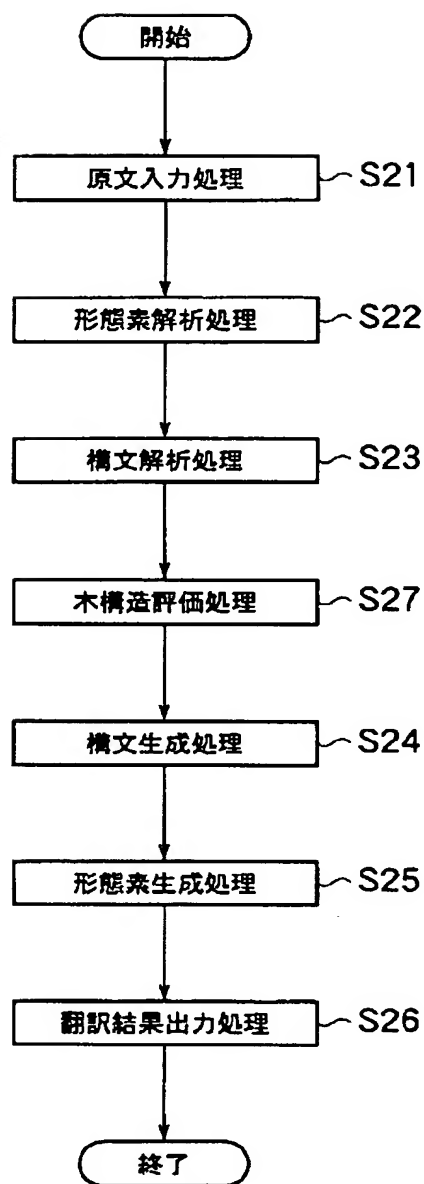
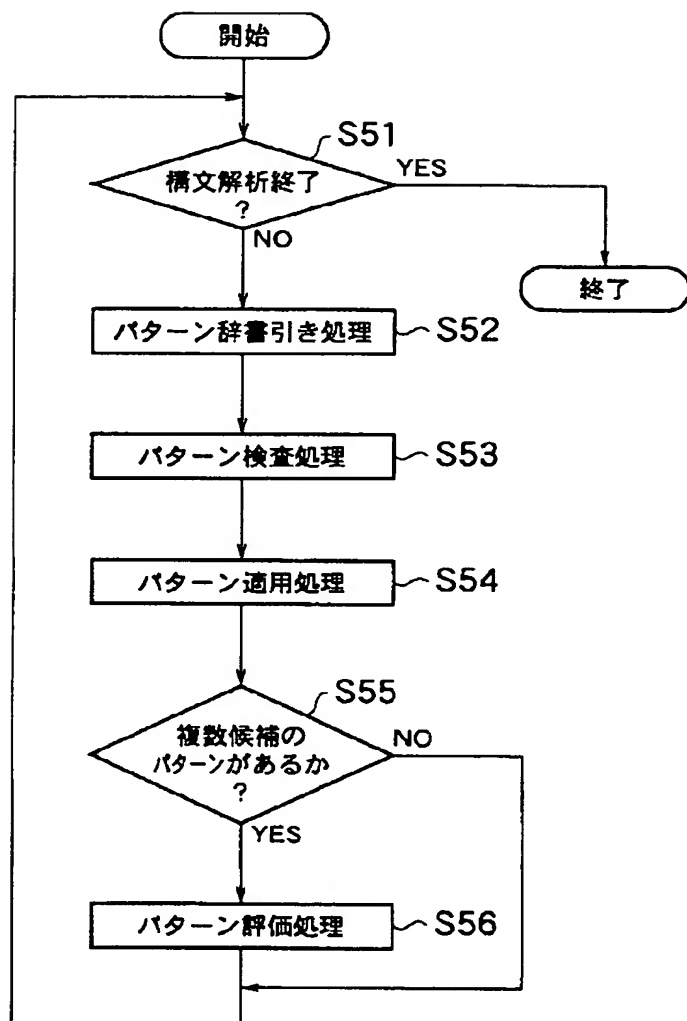
【図22】

	221	222	223
224	user	child	100000000
225	terminal	total	100000000-100000000N
226	i	child	1000000
227	*	total	100000-100000N
228	node	total	1000-1000N
229	!	total	10N
230	+	total	1N

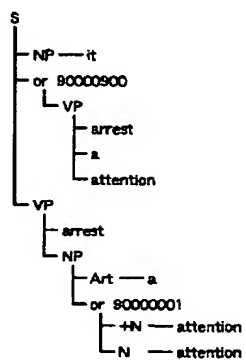
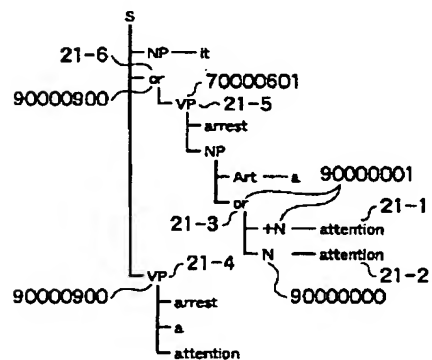
【図10】



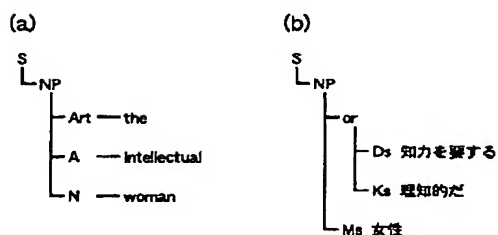
【图 19】



【图 2 3】



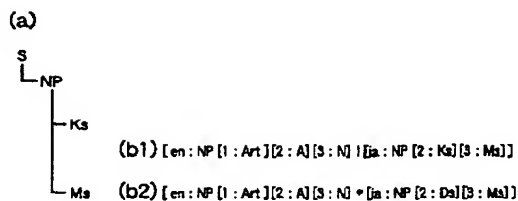
【図 2 6】



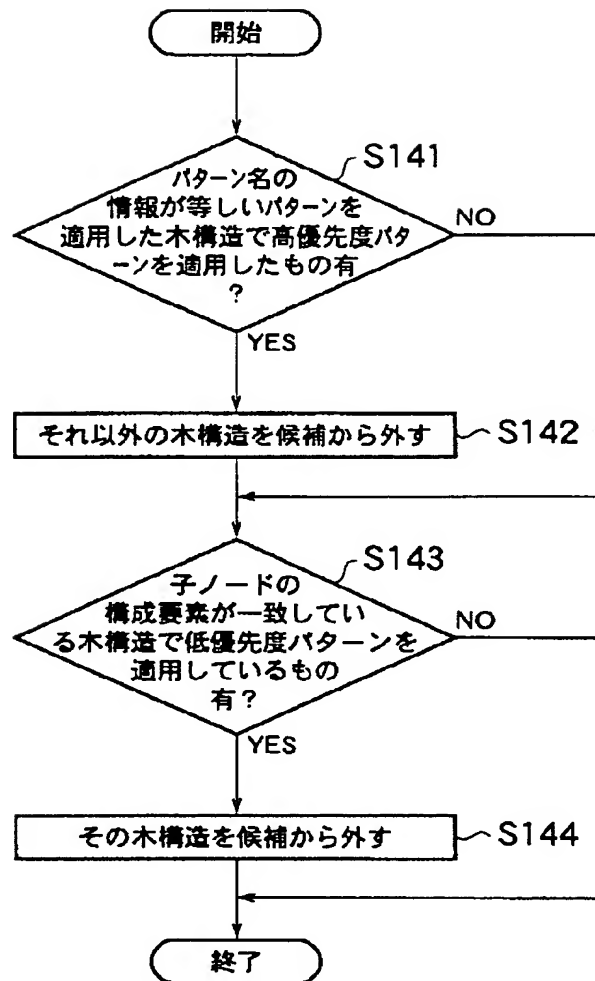
【図12】

12-1	~	[en:VP arrest:品詞 = v [1:NP:意味 = 人]]	[ja:VP [1:NP] を逮捕する]
12-2	~	[en:VP arrest:品詞 = v [1:NP:意味 != 人]]	[ja:VP [1:NP] を阻止する]
12-3	~	[en:VP arrest:品詞 = v a attention]	[ja:VP 興味を引く]
12-4	~	[en:NP:意味 = 物 it:品詞 = prn]	[ja:NP it]
12-5	~	[en:Art the:品詞 = art]	[ja:Art]
12-6	~	[en:N:意味 = 状態 deterioration:品詞 = n]	[ja:Ms 悪化]
12-7	~	[en:N present 品詞 = n]	[ja:Ms 現在]
12-8	~	[en:A present:品詞 = adj]	[ja:Ks 現在の]
12-9	~	! [en:NP:意味 = {1} [1:Art] [2:A] [3:N:意味 = {1}]]	[ja:NP [2:A] [3:N]]
12-10	~	[en:NP:意味 = {1} [1:Art] [2:N] [3:N:意味 = {1}]]	[ja:NP [2:N] [3:N]]
12-11	~	[en:S [1:NP:意味 = {意味}] [2:VP:主格意味 = {意味}]]	[ja:S [1:NP] が [2:VP]]
12-12	~	* [en:S [1:NP] [2:VP]]	[ja:S [1:NP] が [2:VP]]

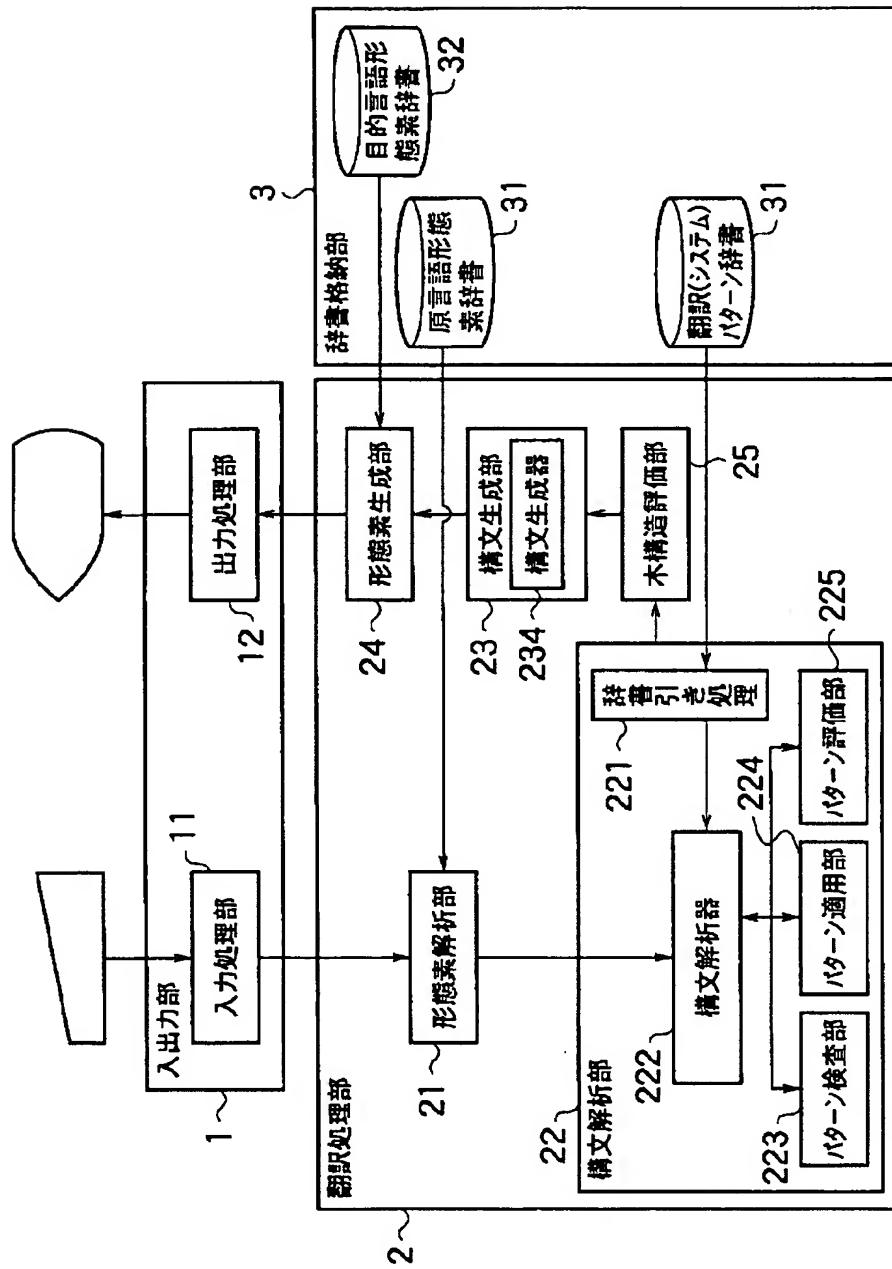
【図27】



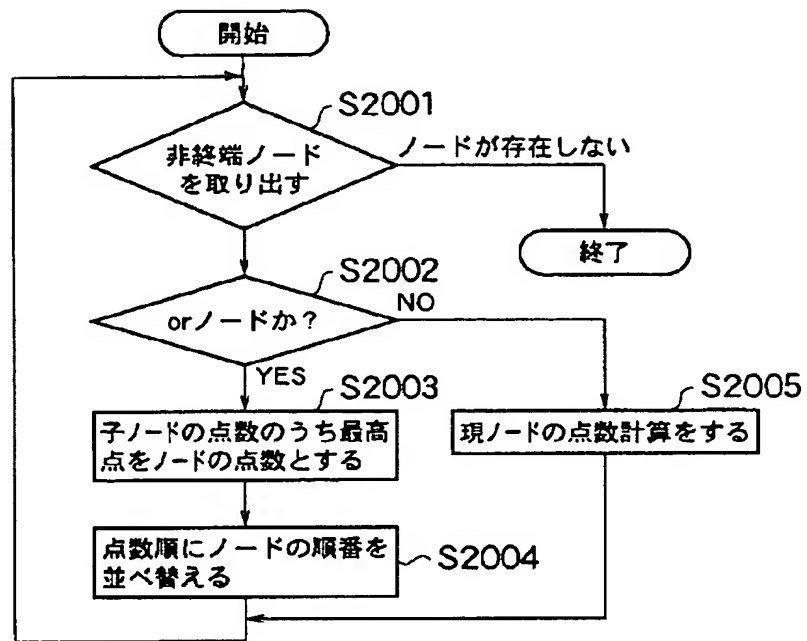
【図14】



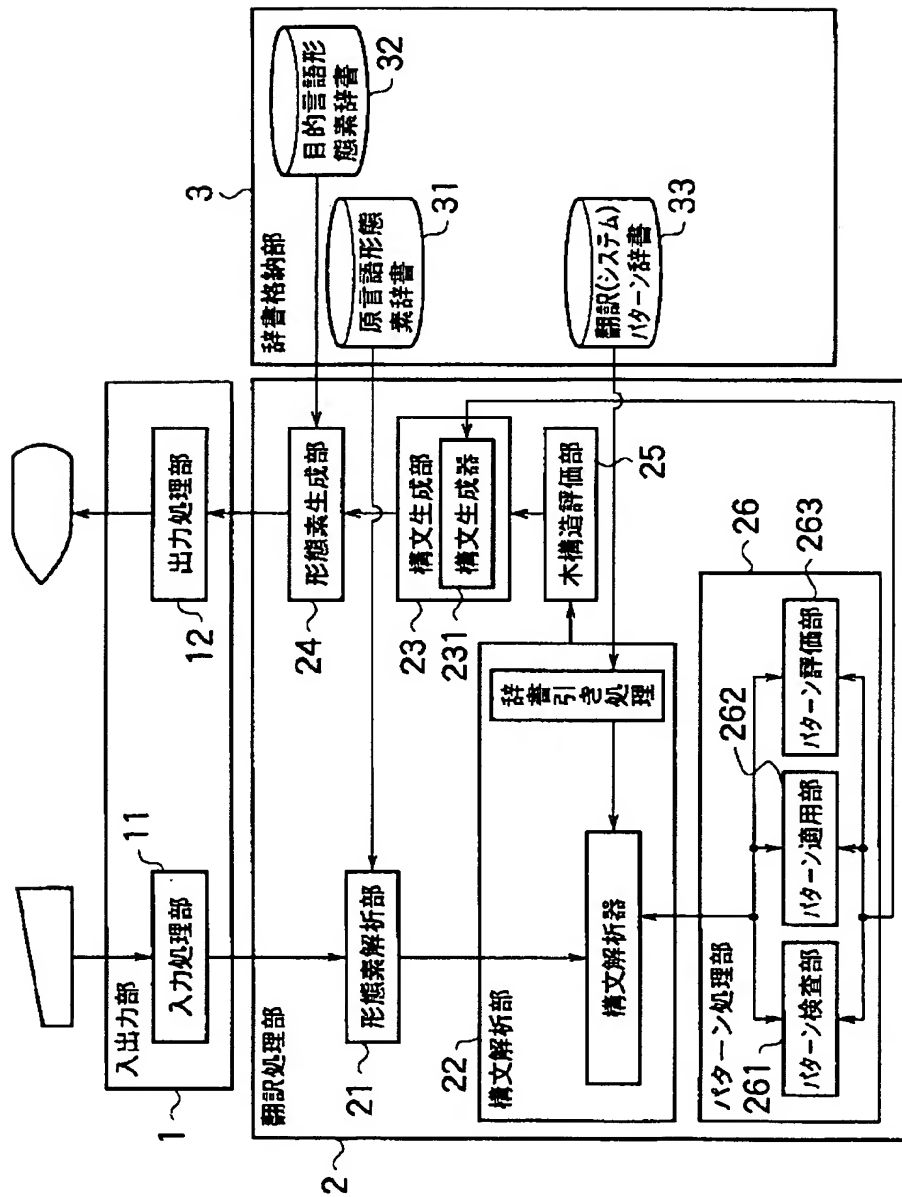
【図18】



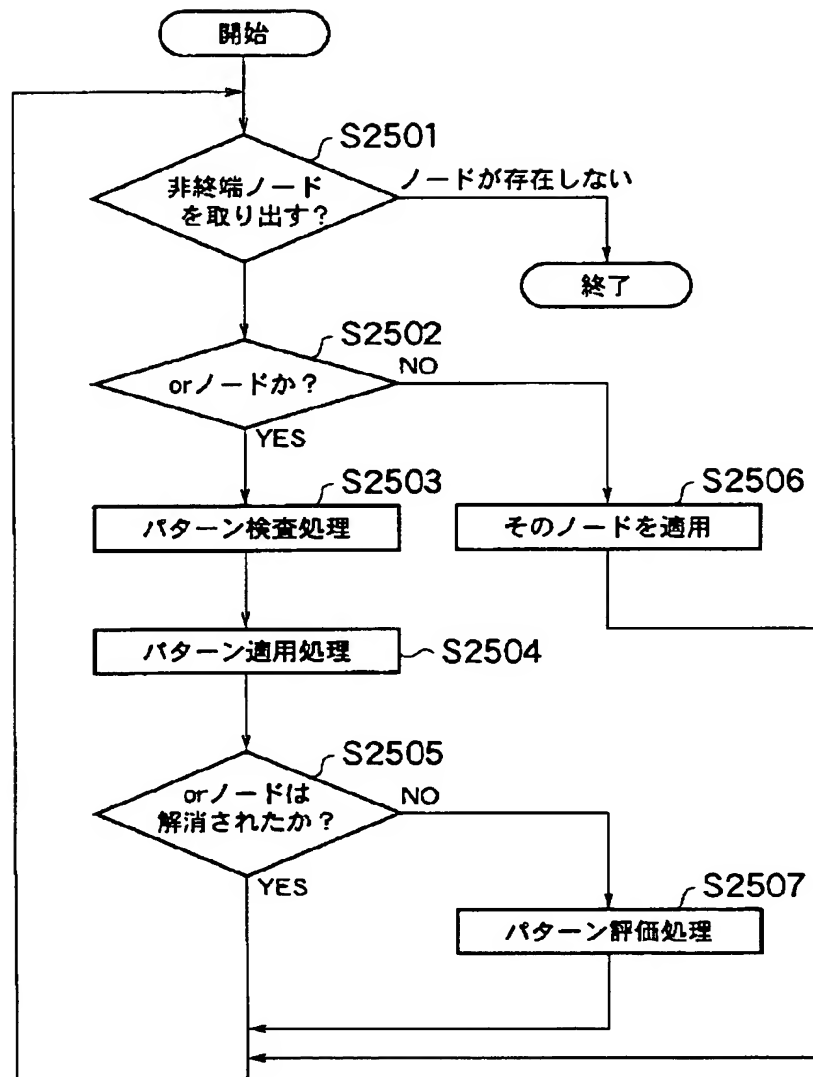
【図20】



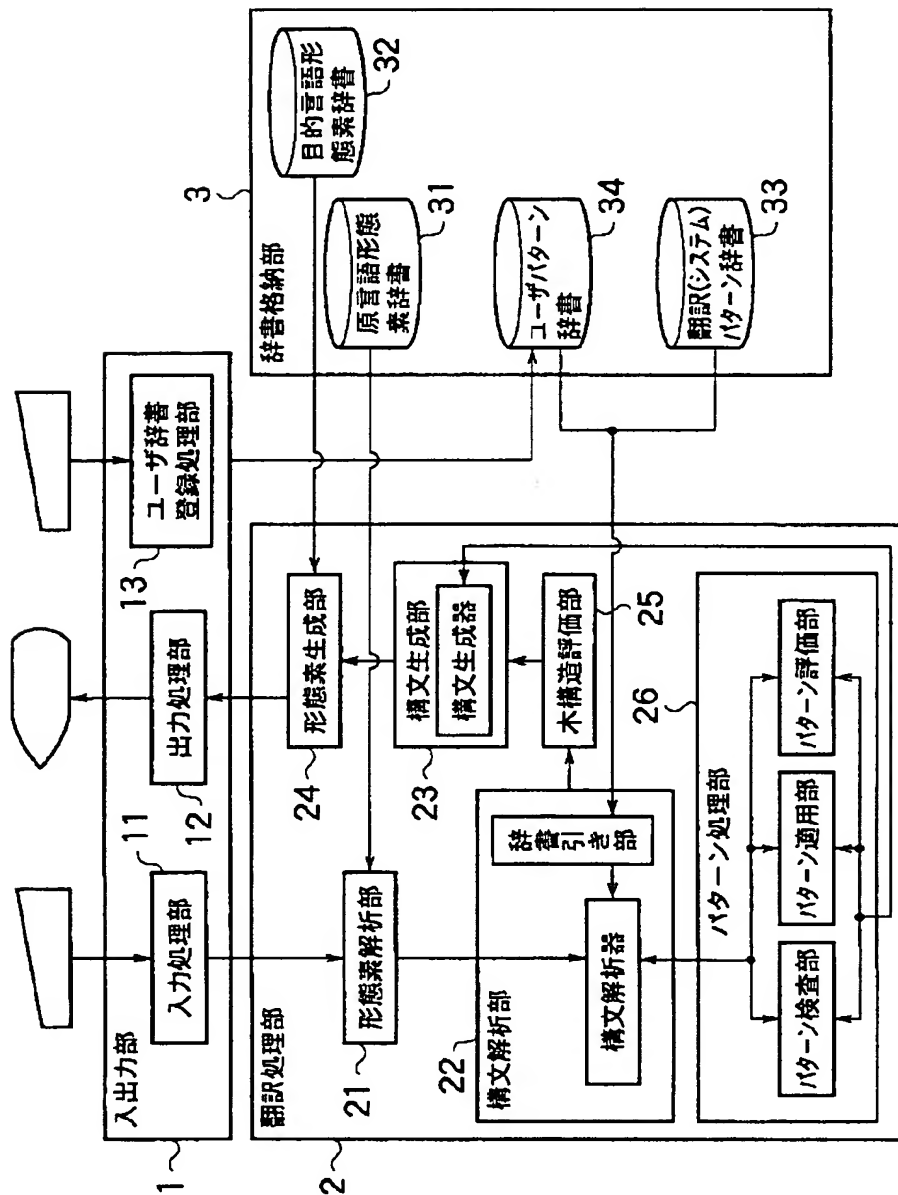
【図24】



【図25】



【図29】



フロントページの続き

(72)発明者 佐々木 美樹
東京都港区虎ノ門1丁目7番12号 沖電気
工業株式会社内

(72)発明者 下畑 さより
東京都港区虎ノ門1丁目7番12号 沖電気
工業株式会社内

(72)発明者 福居 毅至
東京都港区虎ノ門1丁目7番12号 沖電気
工業株式会社内

(72)発明者 淵上 正睦
東京都港区虎ノ門1丁目7番12号 沖電気
工業株式会社内

Fターム(参考) 5B091 AA06 AA15 CA02 CA05 CA24
CC01 CC15